THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Volume XVIII | October, 1927 | Number 7

A COMPARISON OF TWO TYPES OF ILLUSTRATION FOR TEACHING EDUCATIONAL PSYCHOLOGY

GOODWIN B. WATSON

Teachers College, Columbia University

"The trouble with educational psychology is that it isn't practical. It talks too much about white rats, and not enough about the classroom." The possibility that the student who had just spoken these words had hit upon an essential point, was the stimulus for this experiment. Investigation of opinions showed them divided. Some were sure that students needed to learn materials in almost exactly the form in which those materials were to be used; otherwise very little transfer was expected. Others felt equally certain that competent graduate students could make their own applications readily enough, if only the principles were clear.

With the cooperation of Dr. Ralph B. Spence, Instructor in Educational Psychology at Teachers College, this experiment was carried out in the spring session of 1926. Section I, taught by the writer, consisted of about 70 inexperienced undergraduate students of exceptional ability. Their mean intelligence quotient on the Otis Higher Examination was 124. Section II taught by Dr. Spence, consisted of about 30 students, more mature, with previous teaching experience, and with an average IQ on the Otis Higher Examination of 114. Each group had been studying educational psychology for approximately eight weeks at the beginning of the experiment.

The material chosen for teaching consisted of four subsidiary laws of learning as given by Thorndike: the laws of associative shift, multiple response, partial activity, and set. The rotation type of experiment was followed, one section being taught two laws by means of illustrations chosen from teaching situations, and two laws by means of illustrations chosen from non-human relationships. The type of illustration for each law was reversed in the second section.

A preliminary test was prepared to discover the ability of each pupil to analyze certain typical teaching and learning problems. The test contained 12 problems of which the following are samples:

4. A teacher advocating a "direct method" for instruction in foreign language pointed out that ordinary methods were fallacious because the new words were connected only with English words and never with ideas themselves. How far do you agree? Why?

7. A teacher found herself taking a strong dislike to a pupil at first sight and quite without obvious cause. She said she couldn't explain it, but had an "instinctive" feeling that the boy is a sneak. What do you think about it? Why?

9. Suggest conditions under which a 15 year old boy would prefer reading the Encyclopedia Brittanica to eating ice-cream. Explain why.

11. Mr. X teaches history by stating definitely the results of the best available research about causes of historical events. He does this in order to form correct connections in the minds of pupils. Mr. Y criticizes this method, preferring to throw out problems for pupils to think about. "Then when they ask me the answer," says Mr. Y, "I don't tell them." What do you think of these procedures? Why?

The students were given four minutes to write their comments or suggestions on each problem. Notes and jottings only were required, the time being too limited for elaborate composition.

Five days later, no reference having been made to connect the test with this particular day of teaching, the first laws were presented. A uniform method of presentation was followed for each law. The class were asked to try to discover the common principle within the illustrations about to be given. A number of illustrations were presented, the class were asked to describe and name the principle, the instructor brought together the comments into the accepted name and definition, then application of the principle was essayed in several additional situations. It may be characterized as an inductive-deductive method.

Section I was given first the principle of multiple response using the following illustrations:

(*a*) Rat in maze.
(*b*) Cat in box.
(*c*) Marble running down an obstacle board.
(*d*) Brook dammed and finding a channel.
(*e*) Koehler's apes.
(*f*) Hungry dog in closed house.

Next was presented the principle of set, with the following illustrations:

(*a*) Assignment in physics, *vs.* broken radio set.
(*b*) Keeping after school to memorize.

(*c*) "No souls saved after ten o'clock."
(*d*) Classes on excursion see different things.
(*e*) Experiment in spelling words, test by naming in order.
(*f*) Making "pupils' aim" function in lesson plan.

On the same day, Section II studied the law of associative shift, using illustrations such as:

(*a*) Getting child to understand word "pencil" written on blackboard.
(*b*) Acquiring dislike for school because of one failure.
(*c*) Motivation of school work by rewards.
(*d*) Teaching meaning of word "leaf."

In their study of the law of partial activity, the illustrations came from the non-human relationships, such as:

(*a*) Experiments with rats and colored light to indicate path.
(*b*) Koehler's ape using box.
(*c*) Dog following trail.
(*d*) Animal recognizing a disliked individual.

At the next class sessions, two days later, each section was taught the two laws which had been taught to the other section previously, but the type of illustration was reversed, as is indicated by the following samples:

Section II. Multiple response.
(*a*) Teaching child to speak the word "water."
(*b*) Teaching child to write.
(*c*) Behavior trying to get the best of teacher.
(*d*) Acquisition of skill in sewing.
(*e*) Acquisition of skill in manual training.

Section II. Set.
(*a*) Variations in responses of chickens when hungry, well-fed, weak, sleepy, etc.
(*b*) Experimental set in experienced rats.
(*c*) Hungry tiger.
(*d*) Dog looking for its master.

Section I. Associative shift.
(*a*) Pavlov's dog.
(*b*) Teaching dog to sit up and beg at command.
(*c*) Letter diagram showing how response can be shifted.
(*d*) How to teach monkey to run away from a banana.

Section I. Partial activity.
(*a*) Teachers who can quell disorder by lifting an eyebrow.
(*b*) Kilpatrick's story of the boy who refused to say "vertical" when knife was substituted for pencil.
(*c*) Boy who chose books by name of publisher.
(*d*) Learning meaning of triangle, honesty, etc.
(*e*) Need for re-lettering geometry propositions.

Thus each law had been taught by each method, and each section had learned every law, two by one method and two by the other. The next session was devoted to another type of material entirely, and following that came the spring vacation. At the second session following the vacation, the time intervening having been used to start thinking along other lines, the classes were again given the list of problems used in the preliminary test. Not only was no reference made to the teaching of the laws a week before, but a different motive for the re-test was suggested. It was said to be an endeavor to discover the sort of improvement which comes unconsciously after a question has been raised, even when no direct attention has been given to it. With the exception of this explanation, the directions were precisely like those for the initial test. At another session of the classes each group was given ample time to write out from memory "a statement and illustration of each of the four subsidiary laws of learning." This was the first definite reference to them since they had been presented.

It was possible to appraise the results in five ways. From the direct test it was possible to obtain for each person and each law, his ability to remember and state it. These papers were graded by the writer, every answer being graded as right, partially right, or inadequate. From the same test it was possible to obtain in the same way a score for ability to remember and illustrate the law. Noticing a number of illustrations which differed from those used in the presentations, each individual was given a third score for originality of illustration. It later proved that this simple count of original illustrations was not very satisfactory because some of the poorer students, unable to remember what the instructor had said, perforce gave an original although often inadequate illustration, while certain good students having re-thought the material into channels of teaching interest gave new and excellent illustrations. There was a noticeable tendency for original illustrations to come from the teaching field, rather than from the animal or diagrammatic realm.

A fourth score was found in terms of the number of times any particular law was mentioned while trying to explain and to guide the analysis of teaching situations. The fifth score was based upon the ability to handle the situation, to see into it, as one would who was using the law for a guide. It is quite possible, of course, to recognize the element of "set" in a boy's attitude toward ice-cream without using the word or referring to any general law. Likewise the connec-

tions between words and ideas in English and a foreign tongue might be well understood by a student who did not recognize that associative shift was the basis of his analysis.

This fifth score, based upon essay-type papers examined to find the extent to which the pupils made use of the essential insight which a law might be expected to give, might easily have been subjective and inadequate for accurate measurement. Several precautions were therefore taken. An assistant removed the name from each paper and substituted a number, shuffling the papers together so that no unconscious bias toward initial or final papers, toward any section or plan of presentation could operate. A large sample of the papers was read by the two instructors, using a scale giving all the probable responses and assigning complete or half credit to each. This scale was built up from the actual responses found in papers. Two independent readings of 20 papers showed a Spearman correlation of scores given by one and the other of .71. On the basis of this preliminary work a complete questionnaire was prepared, asking specifically for the presence or absence of each of the possible ideas which might be credited. The questions bearing on the illustrations given earlier were:

Case IV.

(A)14. Is it suggested that even though the connection is from foreign word to English, that the English word is connected with the meanings which in the course of time will shift to the foreign word?
Surely. Probably. No.

(P)15. Is it suggested that pupils will notice in the foreign words certain English words which are in part derived from the foreign?
Surely. Probably. No.

(A)16. Is it suggested that meanings for English words or foreign words can grow up through extensive reading in which little by little bad connections are eliminated and relevant ones strengthened by the context? Surely. Probably. No.

Case VII.

(A)27. Is it suggested that the feeling is due to a carry-over from certain unpleasant or sneaky persons previously known?
Surely. Probably. No.

(P)28. Is it suggested that this carry-over is probably due to a few elements such as tone of voice, eye-movements, etc.? Surely. Probably. No.

(S)29. Is it suggested that the teacher may have been over-ready to see it because of a school warning that there were sneaks about, etc.
Surely. Probably. No.

(S)30. Is it suggested that after such an attitude has once been started, many trivial details seem to confirm it? Surely. Probably. No.

(A)31. Is it suggested that the teacher, in order to overcome it, may well notice good points or things which remind her of people she admired? Surely. Probably. No.

Case IX.

(S)37. Is it suggested that some existing interest (*e.g.*, in an argument, a radio, a school assignment) may give a strong set toward the encyclopedia? Surely. Probably. No.

(S)38. Is it suggested that he may be full of ice-cream and have no enthusiasm for it? Surely. Probably. No.

(A)39. Is it suggested that a previous experience with ice-cream may have made him sick, or brought punishment, so he won't want it now? Surely. Probably. No.

Case XI.

(M)45. Is it suggested that because Mr. Y doesn't answer questions, the pupils have to try again in their own thinking? Surely. Probably. No.

(M)46. Is it suggested that when Mr. X points out the right answer, thinking is unnecessary? Surely. Probably. No.

(S)47. Is it suggested that some will grow discouraged if no approval results, and will stop trying? Surely. Probably. No.

The letter in parenthesis in front of each statement was not present on the blank when used for scoring. It indicates that a correct answer for the item is credited on the principle of (A) associative shift, (M) multiple response, (P) partial activity, or (S) set. The total questionnaire recognized the opportunity for 16 A's, 5 M's, 14 P's, and 15 S's.

This questionnaire was filled out for every paper. For most of the papers it was possible to get two different persons to fill out the questionnaire. The agreement, element by element, was a correlation of .33, but the totals used in this study showed an agreement between one set of scorers and the other of over .98.

Table I shows the change which took place on this before-and-after test. The first two columns give the results for the two sections in tendency to make specific reference to the law. It appears that such a tendency was noted in Section I at 3 per cent of the possible points before the teaching, and at 4 per cent afterward. For Section II the preliminary score was 1 per cent, the final score 4 per cent. It appears that there was a slight, very slight gain in tendency to make specific reference to a law. The word "set," probably because it is less technical, appeared more often than did the others. The middle columns represent the ability to analyze teaching situations when measured by the sample marked by instructors in the preliminary reading. The last two give the before-and-after results for all papers,

TABLE I.—CHANGES IN ABILITY TO ANALYZE TEACHING SITUATIONS FOLLOWING TEACHING OF FOUR SUBSIDIARY LAWS OF LEARNING

All Scores Expressed in Per Cent of Total Possible Score, to Nearest Whole Per Cent

	Tendency to explain by specific reference to law				Ability to analyze teaching situations in accord with law (instructor's sample)				Ability to analyze teaching situations in accord with law (all papers)			
Section	I (N = 66)		II (N = 28)		I (N = 37)		II (N = 16)		I (N = 66)		II (N = 28)	
Law	Before	After	Before	After	Before	After	Before	After	Before	After	Before	After
Set	6	9	2	7	24	25	22	21	29	32	24	29
Partial activity	1	2	0	2	15	13	6	11	16	19	12	12
Multiple response	1	2	0	4	20	23	17	31	38	38	22	23
Associative shift	2	3	1	3	13	14	10	13	20	22	16	20
All	3	4	1	4	18	19	14	19	26	28	19	21
Gain ± PE	1 ± 0.2		3 ± 0.3		1 ± 1.6		5 ± 2.9		2 ± 1.0		2 ± 1.8	

studied by means of the complete questionnaire. Section I shows an average improvement of from 18 per cent of a perfect score to 19 per cent of a perfect score when the 37 samples read by instructors are used as the basis of judgment, or from 26 per cent to 28 per cent when all 66 papers are appraised with the more complete questionnaire. In each section and with each method of scoring the gain does not exceed twice its probable error. It appears doubtful whether these students were better able to analyze teaching situations after learning these principles than they were before. The possible improvement may have been simply the influence of a second taking of the test.

This startling ineffectiveness of carefully planned teaching of laws supposed to be fundamental merits serious scrutiny. A possible explanation may be sought in the adequacy of learning. As will be indicated later, in Table II, the groups showed about 50 per cent of perfect score, in ability to define and illustrate. This means, since the items were marked (2) for entirely correct and (1) for partially correct, an average understanding about equal to "partially correct." Some may feel that more than two hours should be spent on these laws. Increase in mastery through further drill might be expected to lead to increased transfer of insight into teaching situations. Such an explanation was checked by selecting the clever students who, on the direct test for ability to recall and illustrate, made perfect scores or lacked only one point of a perfect score. This select group averaged 94 per cent of perfection in ability to define laws and 96 per cent in ability to illustrate. Yet in the analysis of teaching situations embodying these laws they made an initial score of 45 per cent ± 6 per cent (PE) and a final score of 42 per cent ± 5 per cent, or a loss from first to second papers of 3 per cent ± 8 per cent. It thus appears that more efficient learning in terms of ability to define and illustrate laws is no guarantee in itself of improvement in ability to make sound comments on teaching problems.

If it be doubtful whether or not any gain at all has taken place, it becomes doubly difficult to compare types of illustrations. Table II presents, however, the available data. It appears that in ability to remember and state the law, the teaching by means of illustrations drawn from non-human relationships was superior by about 1 PE. In Section I memory was about 60 per cent perfect, in Section II about 40 per cent. In ability to remember and illustrate the law, teaching by teaching situations was superior by about $2\frac{1}{2}$ PE. In tendency toward originality in choice of illustration, those taught from teaching

TABLE II.—COMPARISON OF ILLUSTRATIONS CHOSEN FROM TEACHING SITUATIONS WITH ILLUSTRATIONS CHOSEN FROM NON-HUMAN RELATIONSHIPS

	Ability to remember and state law (score in per cent of maximum possible)		Ability to remember and illustate law (score in per cent of maximum possible)		Tendency to give original illustrations (score in per cent of maximum possible)		Gain in tendency to explain by specific reference to law (score in per cent of all possible opportunities for such reference)		Gain in ability to analyze teaching situations in accord with law: instructor's sample (score in per cent of maximum possible)		Gain in ability to analyze teaching situations in accord with law: all papers (score in per cent of maximum possible)	
Section	I	II	I	II	I	II	I	II	I	II	I	II
Among all papers	59.0±1.8	39.5±2.8	54.0±2.0	43.0±3.0	30±2	47±1	1.7±.2	3.4±.3	0.3±1.6	3.5±2.9	1.7±1.0	2.7±1.8
Teaching illustrations	49.5±2.5	45.5±3.7	60.5±2.8	45.0±4.1	38±1	46±1	1.8±.4	3.1±.4	−0.5±1.9	4.2±2.8	2.7±1.1	2.8±2.2
Non-human illustrations	68.5±2.6	33.5±4.1	47.5±2.9	41.0±4.5	22±1	48±1	1.5±.3	3.7±.5	0.9±1.3	2.7±3.0	.7±2.3	2.6±2.7
Superiority of teaching illustrations	−19.0±3.6	12.0±5.5	13.0±4.0	4.0±6.1	16±1	− 2±1	0.3±.5	−0.6±.6	−1.4±2.3	1.5±4.1	2.0±2.6	0.2±3.4
Resultant general superiority of teaching illustrations	−3.5±3.2		8.5±3.4		7±1		−0.1±.7		0.1±3.0		1.1±3.1	

situations were superior by about 7 PE. It may be that the wider background of experience in teaching situations would make the substitution of new teaching illustrations for those given in class more common than the substitution of new illustrations for those animal experiments which were connected with certain laws. These first three measures were taken from the test on the laws themselves and represent only a final test. It is assumed that the ability to state and illustrate these principles before they were taught was zero, although that it is probably not quite true. The last three columns of Table II deal only with measured gains. In tendency to explain by specific reference to a law, the use of animal and other illustrations was superior by a margin equal only to one-seventh of its PE. In the instructor's samples, read absolutely without knowledge of which were first or last papers, or of section or teaching method used, the teaching situations appeared to give a superior result in analyzing teaching situations, equal to about one-thirtieth of its PE. The most significant measure is that of ability to analyze teaching situations in accord with the law, scored with the complete questionnaire, and based upon all papers. It indicates that the superiority of teaching illustrations is equal to about one-third of its PE. There is thus a slight tendency at every point except the statement of law and reference to law by name, for the method which uses teaching situations to be superior, but the differences are so slight as to be statistically unreliable.

Table III was prepared to answer the question as to the possible superior adaptability of teaching illustrations to the less intelligent members of the class. The results from the two sections were weighted equally and combined, to utilize the advantages of the rotation procedure in eliminating the general ease or difficulty of one pair of laws as contrasted with the other. The high quartile was, of course, generally superior to the low quartile in actual score. It is interesting to note, however, that in the measure of ability to analyze teaching situations in accord with the learning law, the low group gained far more than did the high group. Indeed the top quartile fell off in score between first and second test, on both types of illustration and in both sections. It is difficult to interpret even so slight a loss. Perhaps it means that bright students were bored by the second, rather fatiguing, test. The resultants at the bottom of the table show the extent to which the lower quartile students need more to be taught by teaching illustrations than do the bright pupils. If results are measured by tendency to give original illustrations or by gain in number of specific

TABLE III.—COMPARISON OF HIGH AND LOW QUARTILES (OTIS INTELLIGENCE SCORE) WITH REFERENCE TO EFFECTIVENESS OF ILLUSTRATIONS BASED ON TEACHING SITUATIONS AND THOSE DRAWN FROM NON-HUMAN RELATIONSHIPS

Each Figure Means Per Cent of the Maximum Raw Score, and Is Accompanied by Its PE. Each Score Represents a Combination of Results from Section I With Results from Section II, Equally Weighted

	Ability to remember and state law		Ability to remember and illustrate law		Tendency to give original illustrations		Gain in tendency to explain by specific reference to law		Gain in ability to analyze teaching situations in accord with law	
	High	Low	High	Low	High	Low	High	Low	High	Low
No.	22	16	22	16	22	16	19	17	19	17
Teaching illustrations	60±3	39±5	70±5	54±5	46±3	32±5	0.0±1.3	0.5±0.4	−3.3±1.4	9.5±1.5
Non-human illustrations	64±4	42±4	65±4	32±5	34±6	36±4	−0.7±1.0	−0.3±0.4	−4.2±1.5	4.5±1.5
Superiority of teaching illustrations	−4±5	−3±6	5±6	12±7	12±7	−4±6	0.7±1.6	0.2±0.6	0.9±2.0	5.0±2.1
Superior adaptability of teaching illustrations to the group of low IQ	1±8		7±9		−16±9		−0.5±1.7		4.1±3.0	

references to the law when dealing with teaching situations, then the bright pupils need teaching illustrations slightly more than do the less intelligent. In the other three measures, ability to remember and state law, ability to remember and illustrate law, and gain in ability to analyze teaching situations in accord with the law, the lower quartile students do appear to profit more than bright students, from teaching illustrations in contrast to non-human illustrations. Again the differences are so small that there is barely a 50-50 chance that they would fall in the same direction with other samples.

The following statement of conclusions is not tempered with all of the reservations which can be noted in the above discussion, but is made to stimulate the most effective further thinking and experimentation.

1. If students taught four fundamental and most universally emphasized principles of educational psychology, with an unusual abundance of illustrations in connection with the teaching, show themselves a week later, practically unchanged in their ability to analyze teaching situations, is the teaching of such principles worth while?

2. It is probably true that teaching with illustrations chosen from the field of teaching gives a slightly greater carry-over into the analysis of other teaching situations than is given by animal or diagrammatic illustrations. This may be particularly true of the less intelligent pupils. This difference is at best so doubtful and so slight that anyone who hopes to improve the impact of the teaching of educational psychology upon the professional careers of educators would do well to give his attention to more significant matters. So far as type of illustration is concerned, it need not be exclusively one or the other, but can be both. It seems possible that most texts and teachers use too few illustrations for effective transfer.

3. What is known about the necessity for practicing any mental function in specifically the form in which it is to be used suggests that educational psychology might well experiment with a content and method which will bring much practice in analyzing concrete cases. There may be more promise of building the desired educational insights through such practice than through the teaching of fundamental laws by selected illustrations.

THE EFFECT OF POSITION OF PRINTED RESPONSE WORDS UPON CHILDREN'S ANSWERS TO QUESTIONS IN TWO-RESPONSE TYPES OF TESTS

C. O. MATHEWS

Ohio Wesleyan University

Since the early article by McCall[1] on the use of the true-false test as a type of school examination, many published studies have appeared furnishing data by which this and similar two-response tests have been improved. These tests have been criticised by those who advocate that they fail to give any adequate measure of ability to think. Controversies have arisen around the principle of permitting or even encouraging children to guess the answers to questions unfamiliar to them. Recently an excellent detailed study has been published[2] pointing out some of the factors which influence the validity and reliability of such tests and from carefully analyzed data numerous suggestions are given for their improvement.

One question, however, which seems either to have been ignored or thought of no significance is the possible effect of the position of response words in printed examinations upon the responses made by children. It is this question which has been the center of interest in the present investigation.

Alternate Responses Printed in Vertical Position

Interest in this question grew out of the results obtained in a previous study[3] wherein children were asked to indicate their preferences for numerous samples of reading and graphic materials. These materials were grouped in booklets, each sample being followed by an exercise concluding with this question:

> Do you like this (graph, map, etc.,)
> more or less than (graph, map, etc.,) No. 1?
> Answer by checking one of the words as the right.
> Write no other words..... { more.....
> less.....

[1] McCall, William A.: A New Kind of School Examination. *Journal of Educational Research*, Vol. I, No. 1, January, 1920, pp. 33–46.

[2] Weidemann, Chas. C.: "How to Construct the True-false Examination." Teachers College Contributions to Education, No. 225, 1926.

[3] Mathews, C. O.: "The Grade Placement of Curriculum Materials in the Social Studies." Teachers College Contributions to Education, No. 241, pp. 23–24. This study was financed by a subvention from the Commonweath Fund and carried out under the direction of Professor Harold Rugg.

The booklets were printed at two different times. While one of them was being put on the press the second time the response words *more* and *less* were accidently transposed. These tests were given to children of the same school grades as the correctly printed ones. Since the error was not discovered until the blanks had been returned, no effort was made to exactly match the pupils who had taken the two forms of the booklets. All the blanks containing this question were collected and scored. There proved to be about an equal number with the responses in each of the two orders. All other features of the two blanks were the same.

The total number of responses and the per cent of these which were *more* and *less* for each form of the question are shown in Table I. This table was constructed from the data in Table VIII, page 24 of my study already referred to. It shows the influence of the position of printed response words upon the preferences indicated. The table is read as follows:

TABLE I.—THE NUMBER OF RESPONSES ON EACH FORM OF THE QUESTION AND THE PER CENT OF THESE RESPONSES WHICH ARE OF EACH TYPE

School grade	More / Less form			Less / More form			Difference between columns 3 and 6 or 4 and 7	Probable error of the difference	Difference divided by PE difference
	Total number of responses	Per cent of each type of responses		Total number of responses	Per cent of each type of responses				
		More	Less		More	Less			
1	2	3	4	5	6	7	8	9	10
4	112	85.7	14.3	109	39.4	60.6	46.3[1]	3.9	11.9
5	39	92.3	7.7	42	57.1	42.9	35.2	5.9	6.0
6	49	93.9	6.1	49	38.8	61.2	55.1	5.2	10.6
7	67	82.1	17.9	68	47.1	52.9	35.0	5.1	6.9
8	16	87.5	12.5	17	82.4	17.6	5.1		
9	132	78.8	21.2	134	55.2	44.8	23.6	3.8	6.2
10	34	79.4	20.6	37	48.6	51.4	30.8	7.2	4.3
11	35	74.3	25.7	34	50.0	50.0	24.3	7.6	3.2
12	81	80.4	16.0	82	52.4	47.6	31.6	4.6	6.9
Total or average..	565	83.5	16.5	572	49.7	50.3	33.8	1.7	19.8

[1] The signs used in these differences are all positive indicating that they favor the response printed at the top.

In Grade IV, 112 children answered the $\left\{\begin{matrix}\text{more}\\\text{less}\end{matrix}\right\}$ form of the question. Of these, 85.7 per cent marked *more* and only 14.3 per cent marked *less*. Of the 109 children who responded to the $\left\{\begin{matrix}\text{less}\\\text{more}\end{matrix}\right\}$ form, 39.4 per cent marked *more* and 60.6 per cent *less*. The difference between the per cent who marked *more* on the $\left\{\begin{matrix}\text{more}\\\text{less}\end{matrix}\right\}$ form and the per cent making the same response on the $\left\{\begin{matrix}\text{less}\\\text{more}\end{matrix}\right\}$ form is 46.3. The difference between the two columns showing the per cent of *less* responses would be the same. The probable error[4] of this difference is 3.9, which means that the difference is 11.9 times as large as its probable error.

It is significant to notice that in each grade a larger per cent of pupils marked a given response when it was the upper one of the two alternatives than when it was the lower one. The probable errors of all the differences, excepting Grade VIII, show that they are statistically reliable. (Because of the small number of pupils in Grade VIII and the fact that the PE tables referred to above could not be used for less than 25 cases, the PE was not figured here.) There seems to be a slight tendency for the influence of position to be greater in lower grades since here the differences are slightly larger. This is what one would expect since younger children probably would be less able to discriminate between interesting and uninteresting features of the sample materials used and resort more to chance in indicating their preferences.

Since the two forms of the blank were identical in all other respects and the number of pupils fairly large, the facts in this table may be interpreted as showing that the measure of children's preferences obtained by having the children check one of the two response words printed in a vertical order, such as $\left\{\begin{matrix}\text{more}\\\text{less}\end{matrix}\right\}$, will be greatly influenced by the position of these words. There is a significant tendency for children to check more often the upper one of the two words.

---

[4] The probable errors of per cents and their squares from which the probable errors of the differences were derived in this and following tables were taken from the table by Edgerton and Paterson, *Journal of Applied Psychology*, Sept., 1926, pp. 378–391.

Alternate Responses Printed in Horizontal Position

To what extent does the position of response words printed side by side influence children's answers? The value attached to this method of printing response words is signified by its widespread use. Table II presents a list of the most widely known tests whose response words are printed side by side as *True-false, same-opposite, etc.*

Table II.—Types of Two-response Tests with Printed Response Words Contained in Widely Known Group Measures

Name of test	Order of printed responses				
	True-false	Same-opposite	Yes-no	Strong-weak	No-yes
1	2	3	4	5	6
Army Alpha	x	x			
Bixler's Ten Minute Intelligence Test	..	x			
Brown—Woody Civics Test	..	..	x		
Downey Group Will-temperament Test	x				
Haggerty Reading, Sigma 3	..	..	x		
Haggerty Reading, Sigma 1	..	..	..	..	x
Haggerty Intelligence Examination, Delta 2	..	x	x		
Davis and Hicks Test in Roman History	x				
Illinois Examination, I and II	..	x			
Illinois General Intelligence Scale	..	x			
Michigan Botany Test	..	..	x		
Ohio State University Psychological Test	x	x			
Otis Group Intelligence Scale	x	..	x		
Carnegie Modification, Downey	x				
Stanford Achievement	..	..	x		
Terman Group Test of Mental Ability	x	..	x		
Thorndike Intelligence Examination	x				
Thurstone Psychological Examination	x				
Union Test of Religious Ideas	..	..	x		
Union Test of Ethical Judgment	..	..	x		
Virginia Education Commission General Examination	..	x	x		
Watson Test of Public Opinion	..	..	..	x	
Woodworth Personnel Data Sheet	..	..	x		
Total number of each type	9	7	11	1	1

It is interesting to notice that in every test but one, the true or affirmative response is printed to the left of the false or negative one. There seems to be no published eivdence to account for this coincidence. Perhaps it is due to the tendency for later test makers to follow the examples successfully used in the early group tests such as Army Alpha.

If the position of printed response words helps to influence the number of each kind of responses made, a knowledge of this fact would be useful. First, if this influence were known it could be taken into account in drawing conclusions from the proportion of subjects who mark a given response to an item. Secondly, this knowledge would be useful since any influence which tends to cause subjects to mark one response more often than the other is a factor which influences the validity of the right minus wrong method of scoring. This method assumes that the chances are even that half of the unknown items will be guessed correctly.

Construction of Experimental Tests

In order to determine whether or not the order of printed responses have any influence upon student's reactions, a set of experimental tests was constructed. One hundred questions were drawn up in such form that each could be answered correctly by *Yes* or *No*. These were arranged in order of difficulty on the basis of the writer's judgment and carefully mimeographed in two forms. In one of the forms, the responses Yes and No were printed in this order in parallel columns down the right hand side of the page, each question being followed by the two printed response words. This will henceforth be known as the Yes-no form. The directions to pupils and two samples will illustrate the form.

Draw a line under the right answers as shown in the samples:
Samples: Does a cat have four legs.....Yes No
Are all grapes purple?.....Yes No

The second form was identical with the first in all respects except that the order of the printed response words was reversed, that is, they were printed No and Yes. This will be known as the No-yes form.

The questions were stated so that in every group of ten, five could be correctly answered by Yes, and five by No. No statement was made to the pupils in regard to guessing. It probably would have

been better for the experiment if the students had been instructed to guess since it would have avoided so many omitted items.

Before the tests were given the two forms were alternated in a single pile. That is, the top test was of the Yes-no form, the second the No-yes form, and so on down through the entire pile. This procedure was followed so that the pupils of any one class would be divided about equally as to mental ability, one half taking one form of the test and the other half taking the second form. The time limit for the 100 items was 15 minutes.

Selection of Pupils

In order to make the results on the two test forms independent of the intelligence of the subjects, it was necessary to secure the responses of two groups of pupils equal in mental ability, or to give both forms of the test to the same group of pupils reversing the order of forms in half the cases. The former method was used. In addition to the precaution mentioned above, namely, the alternation of test forms in each room, an intelligence test record was secured for each child.

The graduate students in the major course on intelligence testing at Teachers College, Columbia University, gave either the Haggerty Delta II, or the National Intelligence Test, to each pupil in certain classes in two New York City public schools.[5] These tests were given

Table III.—Comparison of the Yes-no and No-yes Groups of Pupils as to Similarity of Mental Ages

Degree of likeness of MA	Number of pupils according to form of test used	
	Yes-no	No-yes
Number MA's available	16	23
Unable to match MA's within 2 months	27	19
Matched MA's within 2 months	47	47
MA's identical	143	143
Total	233	232

[5] Thanks are due to Louise F. Specht, Principal of P. S. 99, Borough of Queens and to Jennie G. Bowtell, Principal of P. S. 41, Borough of Manhattan, New York City for their kindness in permitting me to use the intelligence test scores and to test the pupils with the experimental tests.

on October 5 and 14 to pupils of the fifth and sixth grades, and to an opportunity class of fourth, fifth, sixth grade pupils. On December 1, the Yes-no form of the test was given to 233 pupils and the No-yes form to 232 pupils in 13 of these same classes. Each pupil tested by the Yes-no form was paired with another pupil of approximately the same mental age who had been tested by the No-yes form. The degree to which this pairing was possible is shown in Table III.

It was decided to use the last two groups of Table III for the final comparison, that is, all pairs of pupils who did not differ in mental ages more than two months. There were 190 such pairs. Their distribution according to school grades is shown in Table IV.

TABLE IV.—DISTRIBUTION OF PUPILS IN EACH GROUP BY GRADES

School grade	Group	
	Yes-no	No-yes
Opportunity 4, 5, and 6	14	13
5 A	38	46
5 B	60	62
6 A	59	52
6 B	19	17
Total	190	190

Comparisons between the two groups according to chronological ages, mental ages, and intelligence quotients are shown in Table V. From these facts and those of Tables III and IV it can be seen that the two groups of pupils are practically equal as to mental ability, so far as this is accurately measured by group tests.

TABLE V.—COMPARISONS OF THE TWO GROUPS OF PUPILS AS TO CHRONOLOGICAL AGES, MENTAL AGES, AND INTELLIGENCE QUOTIENTS

Measure	Chronological ages		Mental ages		Intelligence quotients	
	Yes-no	No-yes	Yes-no	No-yes	Yes-no	No-yes
Range	8–8 to 15–4	8–9 to 14–6	7–9 to 14–10	7–8 to 14–9	61–137	62–146
First quartile	10–4	10– 4	11– 0	11– 0	96.1	96.9
Median	11–4	11– 1	11–10	11–10	105.8	108.6
Third quartile	12–1	11–10	12– 7	12– 9	110.1	116.6

Results of the Experimental Tests

The responses to each question for both forms of the test were tabulated first, as to whether right, wrong, or omitted, and next, as to whether Yes, No or Omitted.

The range in the difficulty of questions is indicated by the fact that the most difficult one was answered correctly by 24 per cent of the children of both groups and the easiest one was answered correctly by 99 per cent of the two groups. No child marked all of the questions correctly. Table VI gives evidence of how well the difficulty of the test was suited to these pupils.

Table VI.—Certain Point and Variability Measures for Each Form of the Test in Terms of Scores[1]

Measure	Form of test	
	Yes-no	No-yes
First quartile	32.88 ± 2.3	31.58 ± 2.3
Median	43.28 ± 2.4	42.64 ± 2.4
Third quartile	54.63 ± 2.4	51.95 ± 2.4
Range	0 to 89	0 to 80
Quartile deviation	10.88 ± 1.6	10.19 ± 1.5
Correlation with MA	.49 ± .055	.68 ± .026

[1] Scores were computed by the rights minus wrongs method.

As has been mentioned the items were printed in what was thought to be an increasing order of difficulty. After the number of right answers were tabulated for each item of the test respective items of both forms were combined and an experimental order of difficulty determined. The correlation, by the product moment method, between the printed order and the experimentally determined order of difficulty proved to be .843 + .019.

The questions were next grouped by tens in order of difficulty, and the number of Yes and No responses for each group totaled. The total number of responses, that is, Yes and No responses added together for each group of items, is shown in Table VII. Column 3 gives these totals for the Yes-no form and Column 6 for the No-yes form, the actual number of Yes and No responses are not shown separately, but were first converted into per cents.

TABLE VII.—THE NUMBER OF RESPONSES ON EACH TEST FORM AND THE PER CENT OF THESE RESPONSES WHICH WERE OF EACH TYPE

Question grouped in order of difficulty	Average per cent of success, both forms	Yes-no form			No-yes form			Difference between columns 4 and 7 or 5 and 8	Probable error of the difference	Difference divided by PE difference
		Total number of responses	Per cent of each type of responses		Total number of responses	Per cent of each type of responses				
			Yes	No		Yes	No			
1	2	3	4	5	6	7	8	9	10	11
1st ten	28.2	1182	49.8	50.2	1132	42.0	58.0	7.8[1]	1.4	5.6
2nd ten	34.9	1320	54.2	45.8	1288	45.9	54.1	8.3	1.3	6.4
3rd ten	41.3	1371	54.9	45.1	1359	47.5	52.5	7.4	1.2	6.2
4th ten	48.5	1485	47.3	52.7	1473	41.5	58.5	5.8	1.2	4.8
5th ten	58.6	1614	51.4	48.6	1617	49.3	50.7	2.1	1.1	1.9
6th ten	66.6	1743	44.6	55.4	1705	40.5	59.5	4.1	1.1	3.7
7th ten	78.1	1781	55.9	44.1	1787	57.2	42.8	−1.3	1.1	1.2
8th ten	84.9	1824	43.6	56.4	1837	42.2	57.8	.4	1.1	.4
9th ten	91.7	1847	49.9	50.1	1857	49.3	50.7	.6	1.1	.5
10th ten	96.2	1868	30.9	69.1	1881	30.4	69.6	.5	1.0	.5
Total or average.	62.9	16035	47.7	52.3	15936	44.5	55.5	3.2	.37	8.6

[1] The signs used in these differences are plus when they favor the response printed to the left and minus when they favor the response printed to the right.

Read the first row across Table VII as follows: To the ten most difficult questions 28.2 per cent of the responses were correct. There were 1182 responses on the Yes-no form of which 49.8 per cent were Yes and 50.2 per cent were No. There were 1132 responses on the No-yes form of which 42.0 per cent were Yes and 58.0 per cent were No. The difference between the per cent of Yes responses on the Yes-No form (49.8) and the per cent of Yes responses on the No-yes form (42.0) is 7.8. The probable error of this difference is 1.4. The difference is 5.6 times as large as its probable error. The differences between the per cents in the columns of No responses and the probable errors of the differences are identical with those between the per cents in the columns of Yes responses.

Table VII shows conclusively that the position of the printed responses in this two response information test affected the answers made by children to the questions. The average difference is 3.2 per cent and the probable error of this difference shows that it could not occur by chance. Furthermore, the difference in every group of questions,

except one, favors the response printed to the left. The reliability of one half of these differences is such as to show that they probably would not be reversed in another similar sample.

Another noticeable fact shown in Table VII is that these differences tend to increase as the difficulty of items increase. The correlation between the differences and the per cent of correct responses on both forms for the ten groups of items is +.88.

If a pupil is absolutely certain of the correct answer to a question, the order in which the response words are printed probably has no influence upon the answer he indicates. But when uncertainty occurs and the pupil begins to take chances by guessing, there is a tendency to choose as the correct response, the one printed to the left more often than the one printed to the right. The more difficult the question the greater is the amount of guessing and the accompanying influence of position of, printed responses.

A wrong response to such a test as the one used may be the result of either of two causes. The child may be misinformed but so absolutely certain of the response he makes that the guessing element does not enter. On the other hand the child may possess *no* information on an item or such a mixture of right and wrong information that the prepotent elements which lead him to guess favor the wrong response. In most tests used with school children the later is probably the principal cause of incorrect responses.

Therefore a tabulation of the *incorrect* responses on each test form and the per cent of these errors which occurred in Yes and No responses, separately, should give further evidence of the influence of the position of printed responses. Such data is given in Table VIII.

It may be noticed in Column 9 of Table VIII, that the differences tend to be about as great for the easier items as for the more difficult ones. The relationships for the ten groups of items is indicated by a correlation of .29. The incorrect responses, regardless of the ease or difficulty of the questions when measured by the average per cent of successes, are very largely the result of guessing. To the child who does not know the answers to an item, even though it is comprehended by a large proportion of pupils, it is difficult and if he guesses at the answer the *position* of the printed response will influence his choice.

The differences in this table are not as reliable as those in Table VII, due to the smaller number of responses upon which they are based. The fact, however, that all but one of the differences favor the response printed to the left is very strong evidence of the existence of a constance influence.

TABLE VIII.—THE NUMBER OF ERRORS ON EACH FORM AND THE PER CENT OF THESE ERRORS WHICH OCCURRED IN EACH TYPE OF RESPONSE

Questions grouped in order of difficulty	Average per cent of success, both forms	Yes-no form			No-yes form			Difference between columns 4 and 7 or 5 and 8	Probable error of the difference	Difference divided by PE difference
		Total number of errors	Per cent of errors in each type of response		Total number of errors	Per cent of errors in each type of response				
			Yes	No		Yes	No			
1	2	3	4	5	6	7	8	9	10	11
1st ten	28.2	590	35.2	64.8	655	29.8	70.2	5.4	3.1	1.7
2nd ten	34.9	605	45.0	55.0	676	37.3	62.7	7.7	2.9	2.7
3rd ten	41.3	551	29.4	70.6	609	23.2	76.8	6.2	3.4	1.8
4th ten	48.5	585	70.1	29.9	530	64.9	35.1	5.2	2.3	2.3
5th ten	58.6	493	31.0	68.0	511	28.6	71.4	2.4	3.5	.7
6th ten	66.6	478	55.9	44.1	439	48.5	51.5	7.4	3.1	2.4
7th ten	78.1	308	37.7	62.3	297	39.4	60.6	−1.7	4.3	.4
8th ten	84.9	214	64.0	26.0	220	58.6	41.4	5.4	4.0	1.4
9th ten	91.7	120	48.3	51.7	101	52.6	57.4	5.7	6.7	.9
10th ten	96.2	48	66.7	33.3	44	61.4	38.6	5.3	8.5	.6
Total or average.	62.9	3992	45.5	54.5	4082	39.4	60.6	6.1	1.1	5.5

What is the probable cause of this influence? This has not been demonstrated by the facts of this study. It is known, however, that on a ballot there are favored places in which candidates long to have their names printed. Children are taught to read from left to right and from the top to the bottom of a page or column, The word which appears at the top or left, in this sense, is the first word. In dealing with facts among which we cannot discriminate, such as a preference for beauty, interest or correctness, perhaps these reading habits cause us to be satisfied more often with the first response suggested.

It has been shown that this influence is greatest where guessing is greatest. In such a test as the Yes and No the child who knows least, probably tends to guess most, at least if definite instructions against guessing are not given. Therefore, in two response-tests with equal numbers of alternate responses the influence of position of responses and the error this would make in a score based upon rights minus wrongs, favors the good student who is influenced only a little by the position of the printed responses.

As was shown in Table VII, the average difference due to position on the Yes and No and No and Yes tests was small, being only 3.2 per cent. If we assume that the correct per cents of Yes and No responses for each item would result from the average of the per cents on the two test forms, then the differences between the average, or true per cent for each group of ten items and the average for either form separately would be one-half of the differences shown in Column 9 of Table VII. One-half of the average difference, shown in the last row, is 1.6. This means that printed position of responses on these tests caused an average error of only 1.6 per cent and the error introduced into a score by the influence of position would be very slight. This error could be avoided entirely, however, by having the position of the printed responses alternated or by permitting the child to indicate his responses by writing a word or symbol. The test maker should take these facts into consideration and at least be aware of the approximate amount of influence position of printed responses is exerting.

If conclusions are to be drawn from the percentage of pupils who respond to any test item by marking printed responses, allowance should be made for the constant tendency of individuals to mark more often the response which is printed above or to the left. This allowance is not necessary as long as pupils or groups are being compared and the per cents are not taken in an absolute sense. But if conclusions should have been made concerning the extent to which one reading passage was preferred more than another from the per cents of preferences indicated on only one form of the $\left\{ \begin{matrix} \text{more} \\ \text{less} \end{matrix} \right\}$ question, a large error would have been made.

Probably the safest way to avoid this error is to let pupils indicate their preferences by writing a response or symbol rather than by marking printed responses.

It is interesting to compare the size of the average difference which indicates the influence of printed position on the $\left\{ \begin{matrix} \text{more} \\ \text{less} \end{matrix} \right\}$ question (33.8 per cent) with that on the Yes-no test (3.2 per cent). The former was a question of preference while the latter was made up of fact or information questions. In the former the alternatives were printed in a vertical position to each other while in the latter they were printed in a horizontal position. It is not known to what extent the larger difference is due to the fact that it is the result of preferences and to what extent it is due to the vertical position of the printed responses.

Summary and Conclusions

1. To a question which involved indicating a preference, there was a constant tendency for pupils to mark a given response more often when it was printed above than when it was printed below an alternate response.

2. The amount of this tendency is indicted by the fact that, on the average a given response was marked 33.8 per cent more often when it was printed above its alternative.

3. To information questions which could be answered by Yes or No there was a constant tendency for pupils to mark a given response more often when it was printed to the left than when it was printed to the right.

4. The amount of this tendency is indicated by the fact that, on the average, a given answer was marked 3.2 per cent more often when it was printed to the left of its alternative.

5. When only errors or incorrect responses were considered, 6.1 per cent more of the pupils marked a given response when it was printed to the left of its alternative.

6. There is a correlation of .88 between the influence of position of printed responses in the Yes-no and No-yes type of test and the difficulty of items, when difficulty is measured by the per cent of successes. There is a correlation of only .29 when the influence of position of printed responses is limited to the incorrect answers.

7. It is suggested that when the percentage of individuals who mark a certain response to a test item by marking printed answers is used as a basis for conclusions as to absolute amounts of knowledge or preference shown, allowance should be made for the effect of the position of the printed responses.

8. In the Yes-no and No-yes test, although the amount of error due to the position of printed response is very slight, it could be avoided in the total score, but not on individual items, by alternating the position of the printed responses. It could be avoided on individual items by permitting the child to indicate his responses with a word or symbol.

AN EXPERIMENT IN MEASURING GROSS MOTOR DEVELOPMENT OF INFANTS AND YOUNG CHILDREN[1]

BESS V. CUNNINGHAM

Institute of Child Welfare Research of Teachers College, Columbia University

The investigation of the motor development of infants and young children is one phase of the periodic examination of a group of slightly more than 100 children ranging in age from 12 to 42 months. The cases were selected in such a way as to make the sampling as near normal as possible. From lists of children living in the neighborhood of the Institute of Child Welfare Research, subjects were chosen at random and parents were invited to bring their children to the Clinic for examination at the time of the birthday and every six months thereafter. Economic and social factors suggested that the group was representative but results to date indicate that it is probably slightly superior in general development in spite of the random method of selection.

The present experimental series of motor tests was devised with the object in view of measuring gross motor coordination as distinct from general intelligence and physical growth as indicated by anthropometric measurements.

In order that verbal directions might be minimized, especially at the one year level it was necessary to arrange situations in which the infant would be motivated to a maximum of effort. This was accomplished by placing toys just beyond his reach and by placing the infant in positions from which he would immediately try to free himself, as for example putting a hoop around his knees. Individual differences in personality reactions to situations in which infants were thwarted and at the same time attracted to objects just out of reach complicated the problem of devising tests of motor ability for the younger levels. The difficulty of verbally presenting the test materials to children of two and three years of age was equally great. Both factors must be considered in a final standardization of the tests.

[1] This is the first report of a series of studies made in the Educational Clinic of the Institute of Child Welfare Research.

The motor tests were devised by the writer in cooperation with the Clinic Committee of the Teachers College Faculty and Dr. Ethel B. Waring and Miss Myrtle McGraw of the Clinic Staff.

Although not devised for the purpose, many test elements could be used in a controlled study of social behavior under emotional stress.

A few of the test situations may be cited somewhat in detail as illustrative of the general procedure. In test one of the series an eight-inch rope hoop is placed around the baby's neck, immediately withdrawn by the examiner and then replaced. The infant practically always attempts to remove it. There are several hoop tests in the series. A flight of three steps with a nine-inch tread is used for a graded series of tests. The infant is placed upon the floor in a seated position facing the steps and with his feet one foot from the riser of the first step. A toy is placed on the floor near the step and is left until the baby throws himself forward and reaches it, or demonstrates either that he cannot or will not take it. If the infant succeeds with this test the toy is placed successively on the first, second and third steps with the subject usually in lusty pursuit. One of the most interesting tests used for the youngest children, and one which calls forth a great variety of emotional responses is that of the long obstacle. The baby is placed upon the floor in a seated position. Between the infant and some toys there is placed a board four feet long, eight and three-quarters inches broad and elevated three and three-fourth inches. To marshal his forces and climb upon the obstacle and down on the opposite side requires more skill than the average infant can command at the age of one year. More difficult tests in the series involve skill in walking, throwing, aiming and balancing.

One of the first steps in evaluating the series is to see to what extent the motor tests are measuring something different from the Kuhlmann revision of the Binet scale. One hundred-thirty-three scores on each test form the basis of the following correlations. There were 89 different children tested. In some instances the same case was

TABLE I.—CORRELATION OF GROSS MOTOR SCORE WITH KUHLMANN MENTAL AGE

Age	Number of cases	r	Number of cases	r
12 months	39	.39 ± .09	23	.65 ± .08
18 months	34	.50 ± .09	23	.45 ± .11
24 months	24	.64 ± .08		
36 months	23	.53 ± .10		

[1] The 23 cases were those selected from the group for a partial correlation discussed later.

included more than once, but at different ages. These and other correlation coefficients are of course to be considered merely as indicative rather than conclusive.

The number of cases is too few for results to be very conclusive but the coefficients suggest a degree of relationship between Binet and motor tests at all levels. Relatively higher agreement might have been expected at 12 months in view of the fact that motor development enters largely into the Kuhlmann scale at this and earlier levels. The necessity for using verbal directions may partially account for the extent of the correlation at 24 months. The relation of the motor tests to tests of abstract intelligence should be made a subject of further study prior to the final standardization of test norms, in order that the motor series may be made to measure a somewhat distinct performance.

Analysis of the motor series in relation to anthropometric measures would seem necessary in further evaluating the motor tests as a somewhat distinct series. For this purpose 122 scores on motor tests were correlated with selected indices. In these correlations a few cases were used twice but at different ages. The results of comparison with estimates of body build based upon the index of $\frac{\text{weight}}{\text{length}^2}$ are given in the following table. These results are not to be construed as an attempt at even a tentative measure of the validity of the motor tests as criteria of motor skill have not yet been determined. Coefficients have been computed with the very small number of cases available in order that the motor test may be considered as somewhat distinct in function from tools of anthropometric measurement now in use.

TABLE II.—CORRELATION OF GROSS MOTOR SCORES[1] WITH $\frac{\text{WEIGHT}}{\text{LENGTH}^2}$

Age	Number of cases	r
12 months	29	.27 ± .12
18 months	28	.15 ± .12
24 months	24	.31 ± .18
30 months	12	−.30 ± .18
36 months	19	−.43 ± .12

[1] From Davenport, C. B.: Height, Weight Index of Build (Tables II and III). Reprint from *American Journal of Physical Anthropology*, Vol. III, No. 4, 1920.

Table II should be studied in connection with Table III, showing correlations between motor scores and $\frac{\text{weight}}{\text{stem}}$.

TABLE III.—CORRELATION OF GROSS MOTOR SCORES WITH $\frac{\text{WEIGHT}}{\text{STEM}}$

Age	Number of cases	r
12 months	29	.45 ± .10
18 months	28	.30 ± .12
24 months	24	.43 ± .11
30 months	12	−.38 ± .17
36 months	19	−.13 ± .15

Considering the two tables together, there is some suggestion of a possible relation between scores on the motor tests and body build. In spite of the fact that the cases are few, the tendency to positive correlation at 12, 18 and 24 months and to negative or zero correlation at 30 and 36 months indicated by both tables may prove suggestive of further study. The significance of relative plumpness, relative length of leg and trunk, relative chest girth and other common measures might well be studied in connection with performance on motor tests. An interesting investigation suggested by the few data so far compiled would be an estimate of motor skill as related to the body build which is somewhat typical of children of given ages.

Assuming that the motor series can test certain skills which have not previously been measured by available standards, the next step in a consideration of the utility of the series is an estimate of the reliability of the measure over a period of six months. Twenty-three cases were available for the following correlation coefficients.

Motor score at 12 months with motor score at 18 months..... $r = .58 \pm .10$
Mental age at 12 months with mental age at 18 months..... $r = .79 \pm .05$
Motor score at 12 months with mental age at 12 months..... $r = .65 \pm .12$
Motor score at 12 months with motor score at 18 months with the mental age at 12 months "kept constant"..... $r = .35$

The reliability of the mental rating of the 23 cases is just about what would be expected with older children. If the correlation of motor score with mental were corrected for attenuation by the Spearman formula, taking the liberty of treating the 12 and 18 months levels as one, the relationship between motor score at 12 and 18 months and Kuhlmann ratings at corresponding levels would be increased to about .82. Even so, there would be indication that there is enough new material in the motor series to continue the problem of devising

a motor scale upon the present basis. When mental age is kept constant the coefficient of correlation is .35. In view of the facts that the comparison was made over a six months interval, and that motor development enters largely into early measures of intelligence, this coefficient seems to indicate that the motor tests may be expected to function satisfactorily as a distinct series.

A further comparison of the predictive value of the test will be available as the 24 and 30 months examinations of the subjects are concluded.

The value of the series may be analyzed further by a consideration the discriminating value of the tests at different age levels. The earlier cases to be examined were selected as a random sampling with the following results.

TABLE IV.—TABLE SHOWING MEDIAN, RANGE AND VARIABILITY OF SCORES AT DIFFERENT AGE LEVELS

	12 months	18 months	24 months	30 months	36 months
Q_3	13	26	37	49	55
M	9	21	29	40	52
Q_1	5	17	26	34	48
Range	2–16	8–35	21–49	25–59	38–62
Number of cases	43	34	30	14	24

The results of the sampling indicate that the test may be expected to discriminate consistently at different age levels.

The next and final preliminary analysis of the tests must consider the content of the series and the type of performance expected at the various ages. For purposes of obtaining first tentative norms in order that certain tests might be retained in the final scale a test is considered applicable at a given age if passed by half or more of the cases. The following scale is based upon a sampling of 128 scores.

This preliminary scale contains the following items arranged in the order of difficulty, the test passed most frequently being listed first.

Twelve Months Tests.

1. To remove a paper cap from head.
2. To walk with help.
3. To stand supporting self.
4. To tap a small bell.

5.5. To remove a hoop from knees.
5.5. To remove feet from box—baby seated on floor.
7.5. To take hoop off of neck.
7.5. To obtain a toy from the second step.
9. To crawl out of a hole in a board elevated 5 inches.

Eighteen Months Tests.

1. To walk without help.
2.5. To obtain a toy from step 3.
2.5. To get off inverted box. Infant seated on box 6 inches high.
4. To get off stool. Infant seated on stool 10 inches high.
5.5. To climb over long obstacle (described above).
5.5. To climb 3 steps.
7. To roll a bowling ball 8 feet, ball weighing 10 lbs., 10 oz.
8. To climb upon low box.
9. To slide or back down 3 steps.

Twenty-four Months Tests.

1. To get off chair, height 13 inches.
2. To climb upon chair, height 17½ inches.
3. To roll a rubber ball half way up an incline, 3 feet, 8 inches long with six-inch elevation.
4. To throw a bean bag into a twelve-inch hole after practice.
5. To roll a bowling ball 9 feet and over a small obstacle.

Thirty Months Tests.

1. To throw a bean bag into hole at 3 feet, twice with three trials.
2. To walk up eight-foot flexible plank elevated 8 inches at upper end.
3. To roll a ball up an inclined board 3 feet, 8 inches long.
4. To walk upon two parallel 4 × 4 × 4 inch beams 6 feet long placed 8 inches apart. (Scored correct if child does not step off.)
5. To walk between straight parallel lines painted on floor 8 inches apart. (Scored correct if child does not step on lines.)

6.5. To step into 3 eight-inch hoops without stepping out or on hoop.
6.5. To roll the bowling ball 10 feet and over obstacle.
8.5. To walk on double diverging beams without stepping off. Beams 4 inches apart at narrow end, 12 inches apart at wider end.
8.5. To walk between converging lines not stepping out more than once, lines 12 inches apart at wide end, 4 inches apart at narrow end.

Thirty-six Months Tests.

1. To roll a rubber ball up an inclined board twice with three trials.
2. To throw a soft ball into a basket from a distance of 3 feet (elevation of basket—3 ft.).

3.5. To walk between converging lines without stepping out.
3.5. To walk on 4 × 4 × 4 beam without stepping off more than twice.
5.5. To walk up steps without support.
5.5. To jump with two feet from eight-inch elevation.
7.5. To walk down steps without support. (Alternating steps not expected.)
7.5. To roll bowling ball 11 feet.
9.5. To throw hoop onto rod from distance of 3 feet, rod elevated 3 feet.
9.5. To start to run within two seconds of signal.

11. To walk on alternating block 3 × 3 × 6 inches without stepping off. (Blocks placed 12 inches apart in each row.)

Tests Too Difficult at Thirty-six Months.

1. To walk on separated blocks without stepping off more than twice. Blocks 3 × 3 × 6 inches placed in straight row 5 inches apart.
2. To run and stop within two seconds of signal.
3. To jump with two feet over hurdle 3½ inches high.
4. To hop on one foot.

More tests are needed at 24 months. Observation of the children so far tested suggests the value of such performances for the two year old child as seating himself in a low chair, picking objects up from the floor, standing on tip toes and reaching.

Summary.—The series of motor tests was devised with the object of measuring gross motor coordination as distinct from general intelligence and general body development.

In order that verbal directions might be minimized the test situations were so arranged as to place the infant in a position calling forth effort.

In response to a thwarting situation many personality differences were evidenced.

The tests for children two years old and older required more verbal direction than will be desirable in final standardization.

Coefficients of correlation between motor scores and Binet suggest a degree of relationship at all levels.

There is some suggestion of a possible relation between score on motor test and body build.

Coefficients of correlation between motor score at 12 months and motor score at 18 months suggest a satisfactory reliability of the motor test.

The test appears to discriminate consistently at different age levels.

The tests scale in difficulty from tests adapted to infants of one year, to children three years old and older.

Suggested Further Steps.—Give the series as situations to elicit responses of emotional variations.

Analyze the relationship between Binet and motor scores aiming to lessen the agreement.

Compare motor scores with numerous physiological indices to study possible influence of body build.

Study the predictive value of the test from 18 to 24 months, 24 to 30 months and 30 months to 36 months.

Standardize the test results with many cases.

AN EXPERIMENTAL STUDY OF VARIOUS GRAPHIC, TABULAR AND TEXTUAL METHODS OF PRESENTING QUANTITATIVE MATERIAL

JOHN NOBLE WASHBURNE

Lincoln School of Teachers College, Columbia University

(*Continued from September Number*)

Numerals and Written Numbers

There is an interesting parallel between the relative standing of written numbers, and round numbers in paragraph form. The two tables in question are forms 7 and 8. The only difference in the two forms is that form 7 contains detailed numbers and form 8 round numbers.[1] The three paragraphs in question are forms 2, 3 and 4. The three forms are identical, except that form 2 contains round numerals (*e.g.*, $5000); form 3 contains mixed numerals and written numbers (*e.g.*, 5 thousand dollars) and form 4 contains entirely written numbers (*e.g.*, five thousand dollars). The meaning and logical arrangement of the numbers is identical in the three paragraphs but the visual pattern is different.

Considering the two statistical tables, forms 7 and 8, the round numbers produced scores[2] considerably higher than the scores produced by the detailed numbers in all questions dealing with specific amounts. In questions dealing with dynamic and static comparisons

[1] Objection may be raised that detailed numbers do not differ from round numbers in detail only, but that the total number is different: In an endeavor to offset this difference the children were told to answer in round numbers or approximate numbers only; and in scoring, greater range in approximation was credited for the detailed than for the round numbers; *e.g.*, one detailed number (10,672,439) was intentionally nearer to the next higher round number (11,000,000) than to the number which appeared in the other forms (10,000,000); for children who had received the detailed number full credit was given for any answer between 10,000,000 and 11,500,000. It is interesting to note, however, that the children who received the detailed number did not give 11,000,000 as the approximation of what they had read any more frequently than the children who had read the round number. In both cases 10,000,000 was the most frequent answer although in the case of the detailed numbers 11,000,000 was actually the more correct approximation. This may indicate that the children responded more to the general appearance of the number than to its inherent meaning.

[2] This phrase should be taken throughout to mean "the children who read the form made scores."

there was not much difference in score. What difference there was seemed to favor the detailed numbers.

Now the superiority of the round numbers in conveying specific amounts might be due either to the difference in meaning of the number (as explained in a foregoing footnote) or to the difference in visual pattern, to one of these two factors being stronger than the other, or to the two factors working in the same direction. An answer to the question as to which of these explanations is probably true may be gained by examining the effect of altering the visual pattern without altering the meaning of the numbers. This is what is done in forms 2, 3 and 4.

If the difference in the two statistical tables is partly due to the visual pattern, what might be expected of the relative standings of the three paragraphs? The difference in the visual patterns of round and detailed numbers is that one is simpler than the other. The simpler pattern (the round numbers) are found to be superior in conveying specific amounts, but not in conveying static or dynamic comparisons. We should expect, then, to find that in the paragraphs also the simpler visual forms are most successful in specific amounts and less so in general comparisons. This expectation is nicely fulfilled.

It is obvious that the numbers in form 2 ($5000, etc.) have the simplest pattern; those in form 4 (five thousand dollars, etc.) have a far more complex visual pattern; while form 3 occupies an intermediate position in the matter of complexity (5 thousand dollars). We accordingly find that in questions dealing with specific amounts form 2 receives the highest score, form 4 the lowest, and form 3 a score about midway between these two scores. In questions calling for static and dynamic comparisons, there is, as in the case of the two tables, very little difference in the scores of the three forms, but what difference there is seems to favor the more complicated patterns.

The conclusion from the foregoing would seem to be that the simpler the visual pattern, the more specific the impression made upon the observer. There is a suggestion that the converse of this statement is also true; namely, the more detailed the pattern, the more general the impression. At least, the fact is strongly pointed out that detailed figures do not tend to produce detailed learning; but on the other hand are fairly successful in conveying general comparisons.

Number and Arrangement of Items

The conclusion that, within the limits of this study, the number of facts to be learned is of much less importance than the arrangement of

those facts, is indicated by the relative standing of all the forms; but it is most clearly shown in the relative standing of forms 6, 8 and 11. The first two of these forms are statistical tables of identical arrangement except that form 6 includes more than twice as many facts as those included in form 8. Yet the latter is not consistently superior to the former. In fact, the table with the larger number of details seems to be somewhat superior in the matter of static and dynamic comparisons, thus bearing out the testimony of the other forms that the greater the specificity of detail the more general the learning. The number of items in form 11, a bar-graph, is less than the number in form 6 and more than the number in form 8. Taking the earnings of a single guild upon a single date as a unit, the three forms in question have 16, 6 and 9 units respectively. Of course, each unit is composed of numerous smaller units.

Now if the dominant factor in learning were the number of items to be learned, one would expect the form with 6 units to produce the highest score, the one with 16 units the lowest score and the one with 9 units a score somewhere between the other two. This however, is not the case; the form with 9 units (the bar-graph) ranks well above the other two in those questions which its arrangement emphasizes; the form with 6 units stands first only in the recall of specific amounts; and the form with 16 units stands first in the matter of dynamic comparisons—a matter in which form 11 is particularly weak and for which form 8 lacks information. However, the table with 16 units is surpassed by the line graph with an equal number of units. And, finally, this same line graph with its 16 units (form 15) outranks, in the matter of dynamic comparisons, all other forms—even those containing only 4 units.

From the foregoing it is plain that if the forms are compared from the standpoint of the number of units they contain, there is very little consistency in the results, except that wherever the *arrangement is constant* the simpler and fewer the units the more specific (and perhaps the more isolated —*i.e.*, less related to other facts learned) the learning will be. But wherever the arrangement is not constant, there is just no consistency to the results. If, on the other hand, the forms are compared from the standpoint of not the number of items but the arrangement of those items, the results are entirely consistent; the line graph is superior in dynamic comparisons, the bar-graph in static comparisons, and the statistical table in specific amounts—no matter how many items are included. Wherever the arrangement is

changed the results change accordingly, regardless of the number of items. We may conclude from this that *arrangement is the dominant factor*, and the number of items the subordinate factor.

The Relative Potency of Logical and Visual Factors

The distinction between "visual" and "logical" factors is given in the "Explanation" under question 3 in the *Outline of conclusions.* Forms 11, 12, 13 and 14 are described there also. It was explained that the logical arrangement of forms 11 and 14 was the same in that both forms grouped the earnings of the guilds according to date (see forms 11, 12, 13 and 14), but that the visual arrangement of the same two forms was very dissimiliar in that one was a bar-graph and the other a pictograph. In 12 and 13 the same facts are listed as in the preceding two forms, but are grouped according to guild. Like the preceding two forms, one of these is a bar-graph and one a pictograph. In short: Forms 11 and 12 (bar-graphs) are alike in visual and unlike in logical arrangement; forms 13 and 14 (pictographs) are also alike in visual and unlike in logical arrangement; forms 11 and 14 are alike in logical and unlike in visual arrangement; and forms 12 and 13 are also alike in logical and unlike in visual arrangement.

By the comparison of these four forms we should be able to answer several questions.

1. Does the logical arrangement have a definite and consistent effect upon the scores independently of the visual pattern?

If the answer to this question is in the affirmative, the scores produced by form 11 should bear the same relationship the scores produced by form 12 as those produced by form 14 bear to form 13: for *in logical arrangement,* 11:12::14:13. But in visual arrangement the four forms are quite differently related. If, then, it is found that the score of each question and group of questions produced[1] by form 11 is related to the score of the same questions produced by form 12 as the scores produced by form 14 (in the same questions) are related to those produced by form 13—in short if all the scores produced by the four forms show the relationship 11:12::14:13—the evidence is fairly convincing, that the same logical arrangement of very different visual patterns produces similar effects.

---

[1] This phrase should be taken throughout to mean "the children who read the form made scores."

The *expected relationship* (11:12::14:13) was *found to obtain in every question and group without exception.*[1]

2. Does the visual pattern have a definite and consistent effect upon the scores, independently of the logical arrangement?

In *visual pattern* 11:14::12:13, hence if the scores of these forms bear the same relationship to each other as the visual patterns bear to each other—that is, if whenever form 11 scores above form 14, form 12 scores above form 13, and whenever form 11 scores below form 14, form 12 scores below form 13—the evidence is fairly clear that similar visual patterns, independently of logical arrangement, tend to produce similar effects.

The relationship 11:14::12:13 *obtained in all groups and in all questions in which there was a significant difference in score.*[2]

3. Since both logical arrangement and visual pattern are of importance, and operate somewhat independently, and since, in graphic representations, they are never found apart, how may their influence be studied separately? And which may be said to be the dominant or controlling factor in the standing of the various forms in static and dynamic comparisons and in specific amounts? That is to say, if one wished to convey the relative earnings of various countries at certain dates (a static comparison) should one endeavor primarily to arrange the facts in a clear and simple pattern, sacrificing any logical grouping which might interfere with such simplicity; or should one seek primarily the appropriate logical grouping, even though such grouping requires a complex visual pattern?

The answer to this question seems to be that for those graphic representations which are intended to impart static comparisons the primary consideration should be logical arrangement, and that simplicity of pattern in this case is of secondary importance. But for those graphs which are intended to impart specific amounts, the reverse is the case. The visual factors (probably simplicity of pattern) are of

[1] Such consistency in this and other *expected* results, gives one more confidence in the unexpected results. It indicates that the groups tested were acceptably equivalent and that the method of scoring was not unreliable. The scores compared were the combined scores of Part I and Part III (a distinction that will be made clear subsequently). But similar results were obtained, by the comparison of umcombined scores. The combined scores are more reliable, and are used whenever all the forms being compared were given in both Part I and Part III.

[2] There were but two questions—1 and 3—in which the relationship did not obtain. These two questions made no marked distinction between *any* of the forms. And in this case there was less than 1 PE difference in score between the two forms which upset the relationship.

primary importance and logical grouping is of secondary importance. For dynamic comparisons the answer is not clear. These conclusions were arrived at by the following comparisons of the four forms, 11, 12, 13 and 14:

Let A designate the logical grouping employed in the bar-graph and pictograph, forms 11 and 14. Let B designate the logical grouping employed in the bar-graph and pictograph, forms 12 and 13. Let X designate the visual factors common to the two bar-graphs, forms 11 and 12. Let Z designate the visual factors common to the two pictographs, forms 13 and 14. Thus form 11 has logical factors A and visual factors X, and may be designated AX; form 12 has logical factors B and visual factors X, and may therefore be designated BX; form 13 has logical factors B and visual factors Z, and is designated BZ; and form 14 has logical factors A and visual factors Z, and is designated AZ.

To summarize:

A = factors common to forms 11 and 14 (logical grouping).
B = factors common to forms 12 and 13 (logical grouping).
X = factors common to forms 11 and 12 (visual factors—bar-graph pattern).
Z = factors common to forms 13 and 14 (visual factors—pictograph pattern).
Form 11 = AX, form 12 = BX, form 13 = BZ, form 14 = AZ.

When the children who read one form make a higher group score in the test than the children who read another form, let the fact be designated by writing one form over the other. For instance, if the children who read form 11 scored higher than the children who read form 13, the fact would be designated by writing the former number over the latter, thus: $\frac{11}{13}$, or by substituting the letters which stand for the forms, thus $\frac{AX}{BZ}$: The four forms under consideration scored as follows in group 8:

Form 11	Form 12	Form 13	Form 14[1]
22.05	28.78	19.79	14.91

[1] These are average percentages of the combined scores of Part I and Part III. They are taken from the special table for the comparison of forms 11, 12, 13 and 14. A special table was necessary because of a misprint in two of the forms which made it necessary to throw out some of the questions. Thus Group 8 is here composed of five instead of eight questions. All the scores used in comparing these four forms are, of course, taken from the same special table.

In this group form 12 scores above 11, 13 and 14. This fact is shown as follows: $\begin{matrix} BX & BX & BX \\ AX & BZ & AZ \end{matrix}$; form 11 scores above 13 and 14, and is written thus: $\begin{matrix} AX & AX \\ BZ & AZ \end{matrix}$; and form 13 scores above form 14, and is written thus $\begin{matrix} BZ \\ AZ \end{matrix}$. In Group 8 (specific amounts) therefore the relationship between the four forms is expressed as follows:

$$\text{Group 8 (specific amounts)} \left\{ \begin{matrix} BX\ BZ & BX\ AX & BX\ AX \\ AX\ AZ & BZ\ AZ & AZ\ BZ \end{matrix} \right.$$

In the first two pairs the visual factors, X and Z, are constant (X is over X, and Z is over Z) and A and B are compared. In both cases B is over A. The conclusion indicated is that, visual factors being equal, logical factors B tend to produce a superior score in specific amounts to logical factors A.

In the second two pairs the logical factors (A and B) are constant and the visual factors (X and Z) compared; in both X is over Z, and the conclusion indicated is that, logical factors being equal, visual factors X tend to produce a higher score in specific amounts than visual factors Z.

In the last two pairs none of the factors is constant, but visual factors X are above Z in both cases. Logical factors B go up when in conjunction with X, and go down when in conjunction with Z. Likewise logical factors A go up when in conjunction with X, and down when in conjunction with Z. This indicates that visual factors X and Z have more to do with the learning of (production of high scores in) specific amounts than have the logical factors A and B.

In Group 6 (static comparisons) there is a relationship which is quite different from the foregoing:

$$\text{Group 6 (static comparisons)} \left\{ \begin{matrix} AX\ AZ & BX\ AX & AX\ AZ \\ BX\ BZ & BZ\ AZ & BZ\ BX \end{matrix} \right.$$

Here, when the visual factors are constant in the first two pairs, A instead of B seems to be the superior logical factor. The second two pairs, in which the logical factors are constant, have the same superior visual factors as in Group 8. In the third two pairs A is above B again,[1] and the visual factors rise or fall according to whether they accompany A or B. Here, then, the finding in Group 6 is reversed in

[1] Note that in Group 8 the first and third two pairs agree as to which are the superior factors; and in Group 6 the second and third two pairs are in agreement. In short, where disagreement was possible there was no disagreement.

two respects: The A logical factors are superior to the B; and the logical factors (A and B) have more to do with determining the scores than the visual factors (X and Z). From this the conclusion was reached that in making static comparisons the logical grouping is of chief importance.

Group 7 (dynamic comparisons) does not offer such clear-cut relationships as the other two groups. This can be accounted for in several ways: First and most important is the fact that three of the scores in this group differ by less than one PE; and, second, the questions which compose this group distinguish less clearly between any of the forms than do the other questions of the test.

Group 7 (dynamic comparisons)	AX BZ	AZ BZ	BZ AZ
	BX AZ	AX BX	AX BX

Visual factors Z are here consistently above X. But when X and Z are constant, A and B do not show any consistent tendency. Hence it cannot be said that these last two pairs indicate the dominance of the visual over the logical factors. It can only be said that the tendency of the visual factors is more consistent than that of the logical factors. However, the strong possibility that further testing would result in a different relation of scores, makes interpretation somewhat superfluous.

The conclusion that the making of static comparisons is influenced chiefly by the logical arrangement of the forms, and that the learning of specific amounts is influenced chiefly by the visual composition of the forms, is indicated by a comparison not alone of the two bar-graphs and two pictographs here under discussion, but also by the forms (2, 3, 4, 7 and 8) discussed above. It was there shown that static and dynamic comparisons were made as readily, or even more readily, by those who had studied complex visual patterns than by those who had studied the simpler ones. But the latter scored consistently higher in specific amounts. This finding is consistent with the cross-comparisons of forms 11, 12, 13 and 14. Such consistency of result, in such different forms, is even more convincing than widely differing scores.

The Effect of Questions Following Graphic Material

Forms 5 and 6 are identical statistical tables except that eight questions concerning it follow the table in form 5. None of these questions deals with specific amounts. Most of them require the making of static comparisons; one of them is a nondescript question which falls into

none of the catagories into which the questions of the test were divided. Some of the questions call for semi-dynamic comparisons (*e.g.*, "How much more did the wool merchants earn in 1175 than they did in 1438?" A genuine dynamic comparison would be: "Between 1175 and 1438 whose earnings increased most rapidly?" There were no questions of this latter variety).

Forms 10 and 11 are bar-graphs which are identical except that form 10 is followed by the same set of questions by which form 5 is followed; that is, by a group of questions concerning static comparisons, a few questions doubtfully related to dynamic comparisons, and no questions at all concerning specific amounts.

What is the effect of these questions? Do they cause the form to be studied more carefully; and thereby increase its effectiveness in general? Or do they change its effect, strengthening it in some respects and weakening it in others?

The answer is clear only to the extent that the scores of both forms —the table and bar-graph—are raised in respect to static comparisons. The scores produced by the bar-graph (the form originally strongest in static comparisons) were somewhat raised throughout. The scores produced by the statistical table were very greatly raised in static comparisons and apparently lowered in other respects. But the loss in the learning of specific amounts was slight in comparison with the gain in the learning of static comparisons.[1] It may safely be said, therefore, that the effect of the questions was, in general, beneficial to both forms.

The Effect of Lengthening the Story upon the Learning of Graphic Material

Do the conclusions which have been drawn in regard to the quantitative paragraph of the story of Florence have any bearing upon quantitative material in textbooks, or would the addition of more reading material and further quantitative material so alter the results as to make the present conclusions worthless?

It was hoped that some indication of the answer to this question might be gained by dividing the story into two parts, each containing a quantitative paragraph, graph, or statistical table and comparing the results of giving these parts separately and in combination.

---

[1] The loss in the learning of dynamic comparisons appeared to be considerable; but for reasons mentioned above these questions are not given much weight.

Part I and Part II of the story of Florence constitute a continuous story with sufficient break in it to make each part seem complete in itself. Each of these two parts contains one graph, paragraph or table dealing with quantitative material. About 50 per cent of the forms of Part I were given alone; *i.e.*, the reading material of Part II was not attached to the reading material of Part I and the Test of Part II was not attached to the test of Part I. The children read Part I, and were tested in Part I; then they read Part II, and were tested in Part II. These tests were scored, tabulated and averaged. Other children read Part I and II together, and took the two tests together. When given together this way the two parts are referred to as Part III.[1] The test of Part III was also scored, tabulated and averaged. Compared with the average scores of Part I it was found to be so similar that the scores of Parts I and III were added together and averaged. It is this final average score which is used in the interpretation of all forms which were given in both Part I and II.

The addition of further reading material containing quantitative data to the material in Part I, (*i.e.*, the addition of Part III to Part I) did not alter the test scores more than the giving of additional forms and tests of Part I without the addition of Part III. In both cases the differences between the scores of the various forms became more consistent. As a matter of fact, the correlation between Part I and the average score of Parts I and III combined is as high as any self-correlation of Part I. For instance, in form 11 (chosen at random) the correlation between Groups 6, 7 and 8 (static and dynamic comparisons and specific amounts) in Part I and the same groups in the combined scores of Parts I and III, using a step interval of 4 per cent, is 1.00. The correlation between all the individual questions in Part I and the combined parts is .98. This is fairly convincing evidence that the

[1] Not all the forms were given in combination. But one form of paragraph, pictograph, line graph, bar-graph and statistical table of Part I was given in conjunction with Part II. All the forms of part II were given in conjunction with form 1 of Part I. This latter because the effect of the test of Part I seemed to be to direct attention so strongly to the quantitative material in Part II, as to reduce very strikingly the correlation between Part II and Part III. To add to the difficulty the material dealt with in Part II was much simpler than that dealt with in Part I, and there was consequently less difference in scores and less reliability. However, in general the findings of Part II supported those of Part I. The statistical table was strongest in specific amounts, the pictograph and bar-graph in static comparisons, etc. Part II consisted chiefly in variations of the statistical table. These variations produced some interesting variations in score. But space does not permit discussion of them in the present summary.

addition of further quantitative and other material to the forms under consideration does not materially alter their effect.

Summary

1. Both logical and visual factors, in the grouping of quantitative material, have a distinct effect upon the recall of the material. Logical factors effect most the recall of relative amounts (static comparisons) and visual factors effect most the recall of specific amounts.

2. The simpler the visual pattern, and the fewer the data, the more specific the recall.

3. The more complex the visual pattern, and the more numerous the data, the more general the recall (and apparently the more effective the making of static and dynamic comparisons).

4. Round numbers (numerals) are more favorable to the recall of specific amounts than are detailed numerals and written numbers. The latter two seem to be somewhat more favorable to the recall of relative amounts and fluctuations (*i.e.*, to the making and recall of static and dynamic comparisons).

5. Increase in the number of data presented in a graph does not effect unfavorably the recall of static and dynamic comparisons, but does effect unfavorably the recall of specific amounts. (In other words, if groups A and B are composed of many complex items appropriately arranged, A can be compared with B as readily as when the groups are composed of a few simple items. But if the detailed contents of the two groups are to be remembered, the number and simplicity of items are important factors.)

6. The paragraph is, in general, the form which is least favorable to the recall of quantitative data, whether general or specific. It is, however, more favorable to the recall of specific amounts than of static and dynamic comparisons.

7. The bar-graph is the form most favorable to the recall of relative amounts (static comparisons) when the comparisons called for involve a fair degree of difficulty. For every simple data some forms of pictograph may be more favorable to the recall of relative amounts than the bar-graph.

8. The line graph is the form most favorable to the recall of relative increase, decrease, and fluctuation (dynamic comparisons).

9. The statistical table is the form most favorable to the recall of specific amounts.

10. Questions following a form increase the effectiveness of the form especially in regard to the recall of those data which the questions concern.

11. The addition of further reading material and quantitative data does not alter the relative effectiveness of the forms.

12. The relative effectiveness of the forms was the same in Grades VII, VIII and IX.[1]

13. There is no correlation between the recall of specific amounts and of static and dynamic comparisons.

14. There is no correlation between the recall of numerical and non-numerical facts.

[1] The scores of Grade IX were considerably higher than those of Grade VIII; and the latter scored higher than Grade VII. But the relative positions of the forms was the same throughout.

THE EFFECT OF EXTENSIVE AND INTENSIVE READING ON THE REPRODUCTION OF IDEAS OR THOUGHT UNITS

CARTER V. GOOD

Miami University

The Problem.—This report is concerned with the effect of extensive and intensive reading on the number of separate ideas or thought units which the reader is able to recall after the reading has been completed. It is possible that certain individuals may acquire from given reading assignments values which necessarily limited tests do not measure. For instance, a given individual may take certain standardized reading tests which require informational answers to a definitely limited number of questions, and have yet other items of knowledge on the subject which the questions do not measure. Therefore, a type of test seems desirable which gives the individual rather complete freedom in expressing the values which are secured from the reading. Certain standardized reading tests measure comprehension of the material read in terms of the number of ideas reproduced or the number of words used in the reproduction. Probably no detailed justification of the use of reproduction of ideas as a measure of comprehension is necessary.

Certain instructors maintain that concentrated study of a relatively small reading assignment in a basal text is highly desirable; other instructors advocate a large and varied reading assignment which the student has time to cover probably no more than once. The experiment here reported was made for the purpose of determining the effect of extensive and intensive reading on a given topic in terms of the ideas or thought units reproduced; it is part of a more comprehensive investigation[1] which measured the values accruing from extensive and intensive reading in terms of other factors such as range and accuracy of information, problem-solving, outlining, and retention.

Method of Procedure.—The investigation involves two types of material, the subject-matter of a given course (reading done without the class period) and material extraneous to the subject-matter of the course (reading done within the regular class period). The first type of material was used in what are known as the course-material tests; the second type of material was used in the extraneous-material tests.

[1] Good, Carter V.: "The Supplementary Reading Assignment." Baltimore: Warwick and York, 1927. Pp. XIV + 228.

The plan followed was to select groups of students which could be equated on the basis of mental ability. The Otis Self-administering Tests of Mental Ability were used to form three pairs of parallel groups. In each instance the average intelligence of the parallel groups was approximately the same. Each group included from 30 to 50 students. It seems that intelligence is the most important single factor in determining comprehension of material read by students at the same academic level. Gray[1] makes the following statements in a recent summary of the investigations relating to reading:

> "With very few exceptions, the studies . . . show a greater or less degree of positive correlation between general intelligence and reading achievement . . . The studies . . . show clearly that progress in reading depends to a considerable extent on the level of intelligence of the learner."

One group did extensive reading on the given topic, while the parallel group concentrated its time on a much smaller reading assignment.

Construction of the Course-material Tests.—For the purpose of the course-material tests two parallel groups in Education 6, "Psychology of High School Subjects," University of Chicago, were used. Two topics of the course were selected for the reading assignments, Generalized Experience, and The Psychology of Study. On the topic, Generalized Experience, the intensive group read only the material (44 pages) included in the text, Judd's *Psychology of High School Subjects;* the extensive group read the material of the text plus 160 pages of supplementary material selected from texts by Thorndike, Colvin, Inglis, and Starch. On the second topic, Psychology of Study, the previous reading procedures of the two groups were reversed, that is, the former extensive group became the intensive group and *vice versa.* On the second topic the intensive group read 75 pages from books by Judd, and Pechstein and McGregor; the extensive group read this same material plus 292 pages of supplementary material selected from books by Burton, Hall-Quest, Miller, Sandwick, and Whipple. For each topic two weeks were allowed in which to complete the assigned reading. The students did not know that a reproduction test would be required of them though they had been told that a test of some kind might be expected.

The chief difficulty in constructing a reproduction test is to make the scoring objective. An idea or thought unit may range in length

[1] Gray, William Scott: Summary of Investigations Relating to Reading. Supplementary Educational Monographs, No. 28, pp. 48, 51. Chicago: Department of Education, University of Chicago, 1925.

from a single word to a long and complicated sentence. The expression of a single thought unit in a complete sentence was accepted as the basis for constructing and scoring the reproduction tests.

The directions which are given below set forth clearly the elements involved in the construction of the course-material tests. This set of directions was used for both the Generalized Experience and The Psychology of Study assignments. The students were directed to state their ideas in separate sentences which were numbered consecutively. This is a device used for the purpose of breaking up the student's knowledge into definite units of expression and to facilitate scoring. Both correct and incorrect examples of ideas on a given topic were listed as guides to the students in the expression of their ideas gained from the reading assignments. The students were cautioned especially with regard to duplicating, paraphrasing, and repeating ideas. A class period of 50 minutes was allowed for the writing, though many of the students did not use the full amount of time.

EDUCATION 6.—GENERALIZED EXPERIENCE

Directions for Reproduction of Ideas

Name.....Group.....Pages.....Time.....

You have been doing some reading on the subject of Generalized Experience. Today you will have an exercise in which you are to write down as many ideas as you recall about this subject. State each idea *briefly*, yet *adequately*, in a sentence and number the statements consecutively. Begin at the left margin of the page with each new idea. Do not duplicate, paraphrase, or repeat ideas. Be sure to make each statement a complete sentence.

Below are some examples of correct ideas on the subject of "Individual Differences."

1. The elective system of the high school is a recognition of the principle of individual differences.
2. The British Royal Commission divides subnormals into four groups.
3. The Binet-Simon test is the most widely used individual test of general intelligence.
4. IQ is based on the relationship of mental age to chronological age.
5. Terman has worked out a revision of the Binet test.
6. Aside from native individual differences, differences may result from environment.
7. The greatest differences in intelligence are found between the white and black races.
8. The heights of a large group of individuals will distribute themselves according to the normal distribution curve.

Below are some examples of statements or ideas on the subject of "Individual Differences" which are too general, incomplete or incorrect to be accepted.

1. There are individual differences.
2. Heredity is one.
3. I think there are many differences.
4. Give the children equal treatment.
5. Retardation exists.
6. Boys are different from girls.
7. Bright children are usually queer.
8. The range of mental ages of children in a given school grade is small.

The examples above were given in order to show you what is wanted on the subject of "Generalized Experience." You will have until the end of the period to finish your writing. At the end of the period fold your papers lengthwise, place your name and group on the outside, and pass to your left.

References read:

1. .....
2. .....
3. .....
4. .....
5. .....

Scoring.—A student's paper is reproduced in full in order to illustrate the method of scoring.

EDUCATION 6.—GENERALIZED EXPERIENCE

Name C. E. S. Group B

POINTS	
1	1. The advocates of formal discipline believe that the training of the mind depends not on what is studied but on the method.
1	2. Those opposed to formal discipline deny that training acquired in a special subject can be carried over to all other phases of mental activity.
1	3. The advocates of each side of the question are bitter against the other side and the controversy has taken a partisan nature.
2	4. It is wrong to assume that no transfer takes place or to assume that habits acquired in a specific activity will carry over equally into other activities.
3	5. The problem in generalized experience is not whether or not there is transfer of training but rather to what degree transfer occurs and by what method.
1	6. Neither party to the controversy has been able to define clearly its position.
1	7. Language is a form of generalization.
1	8. Emotion is another example.

1 9. Generalization is one of the aims of education.
2 10. Transfer occurs only in so far as the two activities have identical elements, either of procedure or substance.
1 11. The classical students are the chief advocates of formal discipline.
1 12. Such experiments as Judd's experiment on target shooting illustrate transfer through generalization.

Student score (points listed separately by student) 12
Corrected score (rescoring by the experimenter) 16

Results of the Course-material Experiment.—Table I gives the scores of the parallel groups of Education 6 on the two course-material tests in reproduction of ideas. Two columns of figures are found under each test. The first column gives the number of ideas according to the numbering of the individual student. The second column lists the corrected scores. The corrected scores are arrived at by the elimination of duplications, the occasional splitting up of a student's sentence into two or more ideas, and by the elimination of wrong or incomplete ideas.

TABLE I.—REPRODUCTION SCORES ON THE COURSE-MATERIAL TESTS

	Test 3		Test 4	
	Student score	Corrected score	Student score	Corrected score
Group A.....	Intensive, 19.5	19.2	Extensive, 17.0	22.4
Group B.....	Extensive, 15.1	17.4	Intensive, 13.9	17.7

The data of Table I show that neither group is consistently superior. The intensive group scores higher on Test 3, Generalized Experience, and the same group as the extensive group scores higher on Test 4, The Psychology of Study.

Such results indicate that for the purpose of reproduction of ideas a relatively small amount of material studied intensively gives at least as good results as a more superficial reading of a much larger body of material. Particularly would this statement seem to be true in the case of a difficult topic such as Generalized Experience. The Psychology of Study is a relatively easy topic; the group which did extensive reading on this topic apparently retained the material with little difficulty and is clearly superior to the intensive group which read a limited amount of material on the subject.

Construction of the Extraneous-material Tests.—It was deemed advisable to construct tests which could be administered within the

regular class period in order that the time element might be controlled. Two sets of material were prepared; the topics were Generalized Experience, and City Zoning. Two parallel groups in education, University of Chicago, read the Generalized Experience material and two parallel groups in social science, University High School, University of Chicago, read the City Zoning material. On each topic there was a condensed treatment of approximately one page which was read by the intensive group. The parallel group read an extensive assignment which included the same page plus approximately two pages of supplementary material. Since both groups used the same amount of time for the reading, the intensive group was able to cover its assignment two or three times. Ten minutes were allowed for the Generalized Experience reading assignment and eight minutes for the City Zoning assignment.

The directions to the students were almost identically the same as those used in the course-material experiment in reproduction of ideas. However, the examples of correct and incorrect ideas were on the subject of Play rather than Generalized Experience in order to be comprehended by high-school students. The preliminary statement which precedes the complete directions is as follows:

Reproduction of Ideas.

The material which is handed you will explain what is to be done. You will be allowed about 15 minutes for the assignment. (Pass out material, explain.) Begin. (Call time at end of 15 minutes and collect papers.)

Fifteen minutes proved sufficient time for all students to express their ideas on City Zoning and many individuals finished early. Twenty minutes were allowed for Generalized Experience since it is more difficult material.

The scoring procedure is the same as that for the course-material tests which is illustrated above by a student's paper on Generalized Experience. To clarify further the scoring technique, the condensed treatment of City Zoning, divided into its thought units or ideas, accompanies this discussion.

CITY ZONING

(1) An important part of city planning is the zoning system. (2) The zoning system controls the height of building, (3) the percentage of lot space which they may occupy, (4) and the use to which they may be put. (5) City planning is entrusted to a commission of public-spirited citizens (6) who usually serve without pay. (7) This commission has the power to divide the city into zones or sections. (8) Each zone is reserved for certain kinds of buildings. Thus (9) one zone may be

for residences, (10) another for commercial purposes, (11) and a third for factories and shops of every description.

(12) The city of Washington was practically forced to adopt the zoning system. (13) The government had once owned the land, (14) but sold much of it to private citizens. (15) Under private ownership speculation took place and (16) values increased enormously. (17) Parts of the city were unimproved, (18) while other sections were very much crowded. (19) "Slum" districts developed in the heart of the city. (20) The large number of government employees who came to the city (21) resulted in the building of many apartment houses. (22) Some were attractive, (23) while others had shops on the first floor and (24) crowded the residential sections. It then became necessary to adopt the zoning system (25) which has improved matters considerably.

(26) A Cleveland judge gives a number of reasons in favor of the zoning system. (27) An apartment or tenement house shuts off the light and air from its neighbors in a section of private residences, (28) invades their privacy (29) spreads soot and smoke throughout the neighborhood, (30) increases the fire hazards, (31) and increases the danger of disease. (32) There is almost constant noise from deliveries. (33) Often the neighboring residences are driven out to make way for more apartment houses (34) and soon the whole street is a row of apartment houses. (35) Apartment houses are a convenience for many people, (36) but private homes are a necessity for people with children (37) who need space for play, grass, trees, and flowers. (38) It is evident that apartment houses should not be allowed in a residential section of private homes.

Results of the Extraneous-material Tests.—Table II gives the results of the extraneous-material tests in reproduction. This table shows that the extensive group scores only 0.1 point higher than the intensive group on the Generalized Experience test.

The data indicate that the intensive group scores nearly 4 points higher than the extensive group on the City Zoning test.

TABLE II.—REPRODUCTION SCORES ON THE EXTRANEOUS-MATERIAL TESTS

Test	Extensive		Intensive	
	Student score	Corrected score	Student score	Corrected score
Generalized experience.....	9.7	8.5	9.4	8.4
Zoning.....	14.1	12.6	15.4	16.5

On the whole, the advantage seems to be in favor of the intensive group so far as the extraneous-material tests in reproduction of ideas are concerned. In general, the answers of the extensive group are more vague and not so clear-cut as those of the intensive group. The extensive group includes some material not found in its reading

assignment and makes relatively numerous repetitions of ideas. In terms of ideas reproduced, the re-reading of a condensed treatment of a given topic is at least as effective as the same amount of time spent in reading an expanded treatment of the same subject.

Comparison of the Test Performances of the Upper and Lower Intelligence Quarters.—Table III compares the reproduction point scores of the upper quarter in intelligence (upper fourth) with the lower quarter in intelligence (lower fourth). A and B are parallel groups; Education 1 and Education 2 are parallel groups; and two classes in modern history are parallel groups. The abbreviations In. and Ex. are used for intensive and extensive, respectively.

TABLE III.—COMPARISON OF THE REPRODUCTION POINT SCORES OF THE UPPER AND LOWER INTELLIGENCE QUARTERS

Quarter	Test 3, A (In.)	Test 4, A (Ex.)	Test 3, B (Ex.)	Test 4, B (In.)	Generalized experience		Zoning	
					Education 1 (Ex.)	Education 2 (In.)	Modern history (Ex.)	Modern history (In.)
Upper.....	21.2	20.5	18.2	13.8	9.3	9.7	14.5	17.7
Lower.....	16.5	21.2	20.2	18.3	7.3	8.0	11.7	15.8

The upper intelligence quarter is not so consistently superior on the reproduction tests as on certain information, problem, and outlining tests which were conducted by the writer in a more comprehensive investigation of which the reproduction tests are a part. However, the average reproduction score of the upper quarter exceeds that of the lower quarter for five of the eight groups.

The suggestion may be offered that students of the lower group in intelligence do relatively good work on a test which involves the reproduction of material that has been memorized in part. It is possible that such a reproduction test does not exercise to the fullest extent the higher thought processes involved in the problem-solving tests, the outlining test, and the exacting questions of the information tests mentioned above. If such an assumption is correct, it may explain in part the relatively effective performance of the lower intelligence quarter on the reproduction tests.

SUMMARY

1. The course-material tests indicate that for the purpose of reproduction of ideas a relatively small amount of material studied

intensively gives virtually as good results as a more superficial reading of a much larger body of material. This statement applies particularly in the case of a difficult topic such as Generalized Experience.

2. The extraneous-material tests indicate that intensive reading is at least equal to extensive reading in terms of the number of ideas or thought units reproduced. The ideas of the extensive group tend to be more general, vague, and repetitious than those of the intensive group.

3. The reproduction point scores of the upper and lower intelligence quarters indicate only a slight advantage in favor of the upper quarter. Such results suggest that students of the lower quarter in intelligence do relatively good work on tests which involve the somewhat mechanical processes of reproduction, recall, and memory. On the other hand, the results of certain experiments conducted by the writer show the much greater superiority of the upper quarter in terms of the higher thought processes involved in problem-solving, outlining, and exacting information tests.

4. If reproduction of ideas is the aim of a particular school subject intensive reading seems relatively effective. However, the question may be raised as to whether other values accruing from extensive reading may not more than offset any limitations it may have in terms of reproduction of ideas.

RELATION OF IQ TO SIZE OF FAMILY

THEODORE LENTZ, JR.

Psychologist, St. Louis Psychiatric Clinic and Assistant Professor of Education, Washington University

That the intelligence quotient as roughly measured by any one of several tests of intelligence has become an increasingly useful guide in the business of individual training appears to be pretty well accepted. Within certain limits, most school people are willing to admit the validity of an intelligence test score for predicting the amount of training which an individual may find profitable and for adjusting the curricula thereto.

While the battle still rages as to the exact amount and nature of the possibilities or limitations so predicted, little is said as to the trend of such possibilities or limitations for the average child of succeeding generations. The constancy of the IQ for the individual has been established by several acceptable studies. This, however, gives us no warrant for assuming constancy of "IQ" for the race or nation over a period of generations; no reason to conclude that the average intelligence of the American people or of the human race will be the same 50 or a 100 years hence. That native human intelligence is a fixed thing, wholly immune to the possibilities alike of becoming better or worse as the decades pass is the generally accepted supposition challenged in this paper, which reports a study attempted for the purpose of throwing some light upon the trend of the population in respect to native intellectual ability. Ascertainment of the trend of intelligence of society as a whole is not necessary merely for the purpose of adjusting the curriculum to any changes which may take place but rather for the purpose of controlling these possible changes in the hope of achieving a higher quality in the raw material which the germ plasm has to offer the educational factory.

Preliminary Statements

During the past two years two-fold data has been gathered by the writer on cases now amounting to over 4000, consisting simply of the intelligence quotient as ascertained by a single intelligence test, and, in each case, the number of brothers and sisters as reported by the subject. This data has been tabulated to show the average IQ of all "only" children as a group, all children of one brother or sister, those

of two, those of three, etc. Transversely, the same data was arranged to show the average size of family of children in the various IQ level groups. Tables I and II present briefly the results of this tabulation.

The cases studied were chosen on the basis of availability from schools in Bridgeport, Conn.; New York City (Brooklyn, Bronx and Manhattan); Englewood, N. J.; Overland, Missouri; Clayton, Missouri; Normandy, Missouri; and Granite City, Illinois. The last four named are all adjacent to St. Louis, Missouri. The children range in age from 6 to 20 and in grade from the first through high school. The median age is about 12 with very few below 9 and very few above 16. The median IQ is 97.9 and the median number of children per family is 4.16.

The IQ's were derived from scores on either the National Intelligence Test, the Illinois Intelligence Test, the Terman Group Test, the McCall Multi-mental Test, the Haggarty Delta I, the Pintner-Cummingham, the I. E. R. Combination or the Stanford Binet. The tests were administered and scored and IQ's calculated by a variety of persons such as research workers and students, principals, teachers and clerks. While it is impossible to compute the accuracy of this part of the study, the writer would place the average on a par with, or above, the group testing in general which is done under non-experimental or practical auspices. The IQ's for the most part were calculated on the 14 year base for those over 14 years of age. The data concerning number of brothers and sisters were obtained in most cases from the notation made by the subject himself at the time of the test or later, no effort being made to check up the accuracy of such reports.

Tables I and II present the basic data of this investigation. It will be noted from Table I that the average IQ of the different family size groups decrease steadily with one exception in passing from the small to the large family children. The 415 children who reported no brothers or sisters have an average IQ of 107.9, whereas the 15 children who reported 12 or more brothers and sisters have an average IQ of 79.9.

Table II presents the same data as Table I but is analyzed by IQ level groups. For example, there are 54 cases with IQ's above 150. These 54 cases had a total of 65 brothers and sisters, making an average of 2.2 children per family. Here again we note the same general tendency as in Table I. Increase in the size of the family is to be noted in passing from the groups of high IQ to the groups of lower IQ.

TABLE I

Number of Brothers and Sisters	Average IQ	Number of Cases Tested
0	107.9	415
1	105.6	865
2	101.5	772
3	97.4	689
4	94.3	516
5	91.8	398
6	88.8	242
7	92.1	181
8	85.5	126
9	84.7	67
10	83.9	25
11	82.6	19
12 and up	79.9	15
Median 4.16		Total 4330

TABLE II

IQ	Average Number of Children in Families Represented	Number of Cases Tested
150 and up	2.2	54
140 to 149	2.5	80
130 to 139	3.0	152
120 to 129	3.1	313
110 to 119	3.5	537
100 to 109	3.5	837
90 to 99	4.2	910
80 to 89	4.9	782
70 to 79	5.3	431
60 to 69	5.4	186
50 to 59	5.7	46
45 to 49	5.5	2
Median 97.9		Total 4330

From both these tables, we gain the impression that there is a marked tendency for the brighter or higher IQ children to come from the smaller families. To present a different numerical picture of the amount of this tendency, and to give an estimate of the amount in different communities, the coefficients of correlation have been calculated between the individual IQ and the number of brothers and sisters.

The coefficients for the different groups and for the composite group are as shown in table on p. 489.

It will be noted that these correlations are low and negative as we would expect from Tables I and II and that the correlations are

Group	Pearson Coefficient of Correlation	
Bridgeport	−.250	±.026
Bronx	−.116	±.029
Manhattan	−.291	±.030
Englewood	−.285	±.031
Normandy	−.265	±.032
Clayton	−.065	±.038
Overland, Grades III, IV	−.273	±.041
Overland, Grades V, VI	−.411	±.033
Overland, Grades VII, VIII	−.321	±.042
Overland high school	−.095	±.053
Granite City	−.264	±.022
Miscellaneous	−.075	±.039
Composite for all cases	−.304	±.010 (4330 cases)

appreciably different in different groups. It will also be noted that the coefficient of correlation for all groups is greater than the average of all the separate correlations. This latter is due to the fact that in some communities the families average large and the IQ is low, while in other communities the reverse is true, and in combining two such communities, the negative correlation between IQ and size of family becomes more apparent. From the point of view of the country as a whole, this composite correlation is the more significnat, for if the communities which have the higher intelligence have also the fewer children, and those which have the lower intelligence have also the more children, surely we have a situation of decline in intelligence that calls for serious consideration.

A study made by Chapman and Wiggins[1] of 1000 cases gives a correlation of $-.33 \pm .024$ between IQ and size of family, while Terman[2] reports a correlation of $-.271 \pm .062$ for some 90 gifted children.

Sources of Error.—It may be well at this point to enumerate some of the possible sources of error in this investigation.

A. Inaccuracy of the IQ's may arise in various ways.

1. Faults inherent in the tests themselves—technically the unreliability of the tests.
2. Faulty administriction of the tests.
3. Errors in scoring.
4. Errors in transmuting to mental age.
5. Errors in chronological age.
6. Errors in computing IQ.

[1] Chapman, J. Crosby and Wiggins, D. M.: Relation of Family Size to Intelligence of Offspring and Socio-economic Status of Family. *Pedagogical Seminary*, Sept., 1925, p. 414.

[2] Terman, L. M.: "Genetic Studies of Genius." Vol. I, p. 117.

B. Errors in reporting number of brothers and sisters.

C. Errors in tabulation.

D. Errors arising from the fact of incomplete families.

E. Errors in sampling. To what extent are these 4000 cases representative of our 20,000,000 school children in the United States?

The amount of the effect of these errors we will not attempt to estimate. The direction of their effect in the case of the first four can safely be assumed. The effect of the first three tend to produce the same result as flipping coins, namely, zero correlation. General errors having no plausible connection with the relationship under discussion, must doubtless lower our correlation. There is no reason to believe, for instance, that in the case of the 415 "only" children, the test results would be erroneously high and that in the case of the 15 children with 12 or more brothers and sisters, the IQ estimated would be erroneously low. Therefore, it seems safe to assume that if these errors were eliminated the correlation would be larger.

Regarding the effect of the fourth source of error mentioned—incomplete families—we can reason more specifically. Of the 415 "only" children (Table I) there are some who may have one brother or sister born after the test was made. This circumstance would place them in the position now occupied by the 865 children having one brother or sister. Judging from what we know of the 865 already in this group we would expect these subsequent additions to have an average IQ of 105.6, or at any rate lower than the average IQ of the "only" children, which average would then be raised by being relieved of these lower IQ's. The average of those in the first group would be further raised by the elimination of those who subsequently have two siblings (or more) and belong with the 772 in the third group whose average IQ is 101.5, or with the 689 in the fourth group, etc. Similarly, the average of those in the second and subsequent groups would be raised by the elimination of those who pass into larger family groups. That is: Withdrawals from any group would tend to raise its average in every case except the lowest and largest family group. This would serve to increase the spread of the groups and could hardly fail to increase our correlation. Further studies could largely eliminate this error by taking into consideration the age of the mother.

The fifth general error is perhaps the most serious as there is no way of arriving at either the amount or direction. Just as we have found a marked variation in the correlation coefficients of the groups from different communities, it would be surprising if we did not find a

somewhat different coefficient for a second 4000 cases selected from other communities in the United States. Obviously, our selection is faulty in not containing any distinctly rural group, where the birth rate is much higher. The Overland groups are from a less urban community than any of the others but still not strictly rural. The only cure for this type of error is to make more surveys and from widely varying communities. The agreement of the composite correlation of this study with that presented by Terman, and that presented by Chapman and Wiggins, previously cited, indicates to some extent that the present selection is fairly representative.

In addition to the facts so far presented, it may not be out of order to enter into a possible interpretation of Tables I and II. Before doing that however, it may be well to raise and discuss briefly, in the light of additional data, the universal question, "Is it nature, or is it nurture?" Do the children from the smaller families receive more attention—more stimulus to mental growth? And may it be possible that the so called test of intelligence is in part measuring the effect of environment so that these differences between large and small families are not to be explained on the basis of heredity but by differences in environmental opportunity?[1] Or on the other hand is it possible that IQ conversely from measuring environment is showing only a fraction of the innate differences which actually exist, and that, were the IQ a true measure of innate ability unmodified by environmental influence, the negative correlation coefficient would be much greater even than the one obtained in this study?

While realizing its insufficiency in settling this general question, data covering a study of educational and achievement quotients along with intelligence quotients made on the Granite City Group (about 700 cases) may be of interest in this connection.

These cases were tested with the Illinois General Examination consisting of tests for intelligence and for reading and arithmetic from which intelligence quotients, educational quotients and accomplishment quotients were computed. The Pearson coefficients of correlation are as follows:

Number of brothers and sisters with IQ	−.264	±.022
Number of brothers and sisters with EQ	−.183	±.024
Number of brothers and sisters with AQ	−.025	±.024

[1] Chapman and Wiggins, in a study previously cited report a correlation of −.27 between family size and social-economic home conditions and likewise a positive correlation of .32 between IQ and good home conditions.

While the negative correlation of EQ with number of brothers and sisters is slightly less than is the correlation of IQ with number of brothers and sisters, it is still negative and marked, while the AQ shows no relation at all with size of family. The difference between the coefficients of −.264 and −.183 may be explained on the basis of the chance variation as indicated by the probable error. Unless we assume that both the EQ and the IQ are enhanced by superior home condition, the absence of any correlation between AQ and size of family would seem to argue against the environmental explanation of the phenomena presented in Tables I and II.

On the assumption that Tables I and II are to be explained on the basis of heredity, that is, that people who have the higher innate intelligence to pass on to their offspring tend for various reasons to limit the size of their families, the following question arises. What will be the effect of such limitation, and of such selection, upon the average intelligence of the general population? To what degree will the average innate intelligence of one generation show decline below the generation preceding. In response to this question, Tables III and IV have been prepared. They are developments, respectively, of the data presented in Tables I and II. Both tables proceed on the assumption that children tend to have the same intelligence as their parents and that the parents will, in any large group where family sizes are equal, have the same average IQ as their offspring. For example, the average IQ of the 415 fathers and 415 mothers of the "only" children in this study, would be approximately 108. This point, taken with the relative number of cases in these family size (an hence IQ) groups furnishes the basis for the computation from which columns 3 and 5 are developed in Table III. For every 415 families with "only" children in the community, there should be 432 families with two children each, and 257 with three. This follows from the fact that the number of children in these groups is respectively 415, 865 and 772, and from the statistical probability that a two-child family is twice as likely and a three-child family three times as likely, to find representation in our group of 4330 cases, as a one-child family.[1] Therefore the number of children in each family size group

[1] Conceive, if you will, of a community made up exclusively of 100 one-child families and 100 ten-child families, making 1100 children in all. Then, out of 110 children chosen at random (without regard to size of family) by the law of chance, there would be 10 children from the one-child families and 100 from the ten-child families. By dividing the number of children from each group by the number of children per family, the relative frequency with which the families of the two sizes exist in the community would be found to be equal.

has been divided by the family size of the group to obtain the relative number of families in each group. Thus, of all families (with children) in the community, 27 per cent (415 out of 1520) would have one child, 28 per cent would have two, 17 per cent three, etc. These 1520 representative pairs of parents would have an average IQ of 102.9 or 4 points higher than that of the children. In other words, the indications are that if all parents had had the same number of children, the raw material with which the public schools now work would be four points higher. Reducing it to mental age, we can say that the average eighth grader, instead of having a mental age of 14-0, let us say, would have one of 14-7.

TABLE III.—SHOWING IQ DECLINE FROM TABLE I

NUMBER OF CASES TESTED	NUMBER OF CHILDREN IN FAMILY	RELATIVE NUMBER OF FAMILIES OF VARIOUS SIZES	AVERAGE IQ	TOTAL IQ PARENTS, VARIOUS FAMILY SIZE GROUPS
415 (divided by)	1 (equals)	415 (multiplied by)	107.9 (equals)	44791
865	2	432.5	105.6	45658
772	3	257.3+	101.5	26157.6+
689	4	172.2+	97.4	16784.2+
516	5	103.2	94.3	9734.4
398	6	66.3+	91.8	6090.6+
242	7	34.5+	88.8	3071.4
181	8	22.6+	92.1	2083.1+
126	9	14	85.5	1196.6+
67	10	6.7	84.7	567.2
25	11	2.2+	83.9	190.6+
19	12	1.5+	82.6	130.8+
8	13	.6+	79.9	47.7+
3	14	.2+		16.1
1	15	.06+		6.2
0	16			
3	17	.01		15.2+
Total		1520.28311		156541.14
Estimated average IQ of 1520 parents				102.9
Average IQ of children tested				98.5
Difference				4.4

A similar difference is noted in Table IV, which is a development of the same notion from Table II. The number of cases tested in each IQ group, divided by the number of children per family for that particular group, gives the relative number of families whose adults have that IQ (on the average). To illustrate: Out of every 1097 families in the community, about 24 would have parents with IQ's of more than 150; 32 with IQ's of between 140 and 149; and 50 between 130 and 139.

TABLE IV.—SHOWING IQ DECLINE FROM TABLE II

IQ	NUMBER OF CASES TESTED		AVERAGE NUMBER CHILDREN IN FAMILIES REPRESENTED		PROPORTIONATE NUMBER OF PARENTS IN COMMUNITY HAVING CHILDREN OF VARIOUS IQ LEVELS
150 and up	54	(divided by)	2.2	(equals)	24.54
140 to 149	80		2.5		32.00
130 to 139	152		3.0		50.67
120 to 129	313		3.1		100.97
110 to 119	537		3.5		153.43
100 to 109	837		3.5		239.14
90 to 99	910		4.2		216.67
80 to 89	782		4.9		159.59
70 to 79	431		5.3		81.32
60 to 69	186		5.4		34.44
50 to 59	46		5.7		8.07
45 to 49	2		5.5		.36
Median 97.9	Total 4330				Total 1097.22

Median IQ for these hypothetical parents	102.0
Median for children tested	97.9
Decline in IQ for one generation	4.1

The median IQ of these 1097 parents is 102.0, which is 4.1 points higher than median IQ of the 4330 children tested. This differs slightly from the variation of 4.4 found in Table III. The latter, however, is figured on the basis of the mean, whereas Table IV shows differences between medians.

Any serious consideration of these differences would include the question of error in the data or its treatment. It is hoped that publication of Tables III and IV will challenge thinking sufficiently that possible errors in the process of reasoning will be pointed out. For the purpose of this paper it will suffice to say that the errors previously suggested apply here as well as to the original data in Tables I and II. Errors due to inaccurate report of family size, faulty tabulation and incomplete families all work in one direction, namely, to reduce the correlation, and, in Tables III and IV, cause the differences between the two generations to appear less than it really is. The same tendency would in part result from inaccuracy of IQ. This error however works to unduly accentuate the real difference through the effect on spread. The extremes of Table IV, for example, appear to exercise a very great influence upon the difference. Another source of error previously mentioned, namely that due to sampling, is indeterminable as to amount and direction. In order to get a check on the sampling here found, a comparison has been made with the census

reports of 1923 on the number of births by order of birth, for the registration area. By subtracting the number of second born from the number of first born we get a rough idea of the number of mothers who may be expected to stop with one child. Likewise, by subtracting the number of third born from the number of second born, we have a rough indication of the number who will stop with two children etc. The results of this calculation differs greatly from column 3 of Table III in being more heavily weighted with parents with larger families. Substituting these figures for those in Table III, column 3, the difference between the two generations appears much greater than four points.

In further attempting to ascertain the representative nature of our sampling, the question of normal distribution of IQ's has arisen. Table V compares the distribution of IQ's, (1) according to Terman,[1] (2) according to Army Data,[2] and (3) according to Tables II or IV of this study.

TABLE V

IQ	IQ Percentages According to Terman, Per Cent	IQ Percentages According to Army Data on Principal Sampling of 93,955 Cases from White Draft, Per Cent	IQ Percentages 4,330 Cases of This Study, Per Cent
Below 70	1	6.2	5.4
70– 79	5	8.4	10.0
80– 89	14	14.4	18.1
90– 99	30	21.5	21.0
100–109	30	19.4	19.3
110–119	14	14.2	12.4
120–129	5	8.8	7.2
Above 130	1	7.1	6.6

It will be seen from Table V that the distribution of cases in this study and that of the 93,955 cases cited from the army tests agree fairly well with each other though both show a wider spread than that of Terman. This is probably due in part to the effect of group testing, since Terman's figures are for individual tests. If we use Terman's

[1] Terman, Lewis M.: "Measurement of Intelligence." P. 78.

[2] Memoirs National Academy of Sciences, Vol. XV, p. 790, Table 333. The percentages shown in column 3, Table V, are on the basis of the combined scale—Army Alpha and Army Beta. The IQ percentages are calculated from Table 333 of the Memoirs, which lists the percentages at various mental age levels. The percentages here shown are for IQ's calculated on 13.2 as the chronological age basis. in order to give symmetry to the distribution.

percentile distribution and the figures from column 2 of Table IV we get a markedly smaller difference between the two generations.

A further source of error from the standpoint of Tables III and IV lies in the fact that we are dealing with only a part of the generation to which the parents belong. Childless persons are not represented in this study. Judging from the trend in Table I they would appear at the head of the list with high IQ's which in turn would cause the difference between the two generations to be greater than here appears.

Summary

A study of 4330 cases shows a marked inverse relation between size of family and IQ as determined by group tests of intelligence. This negative correlation differs with groups representing different communities and tested with different tests.

The inverse correlation between IQ and size of family is sufficient to indicate a marked decline in average IQ from one generation to the next. The amount of this decline cannot be adequately measured until more studies of a similar nature on representative groups are made, and some account taken of the childless persons. A zero correlation between the number of brothers and sisters and achievement quotient suggests but does not prove that home environment is not responsible for the superior IQ of children from the smaller families.

If, as these figures indicate (along with other indications that may be noted), the sum total of our intellectual assets is diminishing, whither shall we look for the remedy? Certainly no increase in mere number of children born can compensate for the decrease in brightness. A popular street advertisement once read: "Its doing better than the average that keeps the average up." To whom shall we look to better the average of our IQ? There are, broadly speaking, two groups of people who can help: Those above the median, and those below it; the former by increasing their fecundity, the latter by diminishing it.

A COMPARATIVE STUDY OF THE EDUCATIONAL ATTAINMENT OF NEGRO AND WHITE CHILDREN

PAUL A. WITTY

University of Kansas

AND

A. I. DECKER

Superintendent of Schools, Coffeyville, Kansas

The inadequacies of our present devices in educating the negro are recognized by most careful educators. Materials and techniques of instruction which produce satisfactory results with white children appear to be poorly adapted to the needs of the negro. The problem of segregation, too, is a crucial one.

In many school systems segregation of the races is practiced in the grades. However, in the high school negroes and whites are often enrolled in the same classes. Here the restricted educational attainment of the negro often presents an insurmountable burden when group instruction is attempted.

The inefficiency of our present system has been recognized by many superintendents and principals. Yet, there has been little scientific experimentation to determine accurately the educational status of the negro at various periods in the elementary school. Then, too, there has been little experimentation in evaluating methods and material of instruction for the negro.

The starting point for remedial work and further experimentation involves an accurate knowledge of the educational status of representative negro pupils enrolled in the elementary schools. The following paper reports a comparative study of the educational attainment of negro and white children selected from the elementary schools of Coffeyville, Kansas.

METHOD

The Stanford achievement test was given to 1725 white and 220 negro pupils of chronological ages 7 to 13 inclusive. Children of these ages were selected in order that a representative sampling of each race might be obtained. Children over 13 were eliminated from the study since among these children the number of overage negro children far exceeded the number of overage white children. A fairly representative sampling of each race was secured among children 7 to 13 years of age. The 220 negro pupils chosen were distributed fairly uniformly in each year interval from 7 to 13.

The Stanford achievement test measures attainment in the following subjects: reading, arithmetic, nature study and science, history and literature, language usage and spelling.

The directions for giving and scoring are all in a single manual. The norms for the different tests are similarly derived; this fact renders a pupil's score in any subject immediately comparable with his score in any other subject. Age norms as well as grade standards are available for all the tests, thus making possible comparisons of the attainment of individuals or groups with average attainment of public school pupils of like ages or grades.

The time required to give the entire examination is about three hours. The time allowances for all tests are liberal enough to make the test almost entirely a test of *power* rather than of *speed*.

The reliability coefficients of the tests for the different subjects are extraordinarily high. The coefficient of reliability for the entire battery of tests is +.98 for unselected pupils of a given age.

The tests were administered in a single week and uniform procedure was followed in giving and scoring the examination.

Results

The composite score upon the entire examination and its equivalent educational age were obtained for each child.

Table I shows the number of months that the median educational age of white children exceeds the median educational age of negro children at each age level.

Table I.—Average Number of Months White Children Exceed Negro Children in Educational Age at Each Chronological Age Level

Chronological Age	Educational Age Difference in Months
7	8
8	14.3
9	14
10	17
11	17
12	25.3
13	27.4

It is apparent from Table I that white children excel negro children in educational age in every age level. Particularly salient is the fact that younger negro children reach more nearly the educational status of white children than older ones.

The number of months by which the median educational age of the entire group of white children exceeds the median of the negro

group was found to be 16.7 months. It was found further that *only 14½ per cent of the negro children reach or exceed the median educational age of the white children.*

Data were next assembled for the races upon each part of the test. Table II shows the number of months that the white pupils exceed negro pupils in median subject age in each section of the test.

TABLE II.—AVERAGE SUBJECT AGE DIFFERENCE

	Reading	Language usage	Nature study and science	Spelling	Arithmetic	History and literature
Difference in Educational Age (months).	20	17.4	16.6	15.6	14.4	10

From this table it is clear that negro children nearest approach the educational status of white children in history and literature. Next in order are arithmetic, spelling, nature study and science, language usage, and reading. The two subjects in which negro children manifest *poorest scholarship* are reading and language usage.

A further examination of the test data revealed the following facts:

1. White children excel negro children in median subject age in *every test at every chronological age level.*

2. In the three parts of the reading test negro children make their best scores in word meaning, next in sentence meaning and do poorest in paragraph meaning.

3. There is little difference in the educational ability of negro children in the two parts of the arithmetic test. Arithmetic reasoning and arithmetic computation scores yield subject ages that differ little.

4. Younger negro children reach more nearly than older ones the educational status of white children in *every section of the test.*

REMARKS

This study substantiates previous studies in demonstrating the educational inferiority of the negro race. The relative educational retardation may be due to racial inferiority in mental ability. There are several conditions revealed by the test data which make one hesitant to accept mental inferiority as the dominant cause of the educational retardation. These conditions are:

1. Younger negro children approach more nearly the standards reached by white children than do older ones. There is some evidence

showing that the intelligence quotient is relatively constant. If this hypothesis stands, there is no reason to assume that mental inferiority is a more potent factor in effecting poor academic attainment in years 12 to 13 than in years 7 to 8. The writers attempted to secure representative negro children at each age level.

2. Negro children approach the standards for white children more nearly in the history and literature test than in any of the other tests. It seems reasonable that *mental ability* is a potent pre-requisite for success in these subjects. The success of the children upon this test suggests that the negroes studied must be functioning far *below capacity in many school subjects.*

3. Negro children evidence poorest scholarship in reading and language usage. Mastery of these subjects is essential for success in most academic activity.

One must turn, therefore, to sources other than mental inferiority for the explanation of the educational status of the negro.

The fact that younger negroes attain more nearly than older ones educational status of white children may be explained in terms of difference in the maturation of the two races. It may be that intelligence matures more quickly in the case of the negro than in the case of the white child. Experimentation is needed in this regard before exact statement can be made.

There is another factor which must be considered as a possible one in effecting the ever widening difference between the two groups as chronological age increases. There is evidence that equal amounts of practice may increase the difference between two groups when the groups differ widely in ability in the initial stage.[1] Certainly, administration of the same course of study to the two groups has resulted in widely varying results. Effort was made in the Coffeyville Schools to administer the course of study impartially to the two races.

No attempt is made by the writers to state dogmatically the causes for the educational inferiority of the negro. The hypotheses stated above are offered merely as suggestions.

The writers feel that the data herein presented give evidence of conspicuous needs. Subject-matter and methods used in the education of negroes should be examined carefully. Experimentation should be directed toward developing more effective materials and techniques of instruction. For the negro group, it is particularly vital that remedial work be provided by which reading and the language arts may be taught more effectively.

[1] Thorndike, E. L.: *Educational Psychology*, Vol. III, p. 305.

NOTES ON ARTICLES IN EDUCATIONAL PSYCHOLOGY IN CURRENT ISSUES OF OTHER MAGAZINES

REPORTED BY C. O. MATHEWS
Research Associate, Institute of Educational Research, Teachers College
Columbia University

Intelligence Testing

Prediction of Reliability of Mental Tests and Tests of Special Abilities. Lyle H. Lanier, Journal of Experimental Psychology, April, 1927, 69–113. From experimental data the author concludes that use of the Spearman prophecy formula to reliabilities which would occur by the increase of time or number of subjects is not justified in the sense that many workers are using it.

Sex Differences in Army Alpha Scores in the Secondary School. Guy M. Whipple. Journal of Educational Research, April, 1927, 269–275. The author concludes from data given that boys as a group are slightly superior to girls. This may be due to the selective influence of the high school.

Clinical Values of the Rhode Island Intelligence Test and the Town Picture Game. Louise E. Poull. The Journal of Applied Psychology, February, 1927, 68–72. These two tests are recommended as useful instruments for clinical cases on the basis of their use with 100 cases for the Rhode Island test and 69 cases for the picture game.

Group Intelligence Tests and Linguistic Disability among Italian Children. Margaret Mead. School and Society, April 16, 1927, 456–468. When the group was subdivided into four groups on basis of amount of Italian spoken in the home there was found to be a steadily increasing mean on the Otis Group Intelligence Test from groups speaking only Italian to those speaking only English.

The Transfer Effect of Practice in Related Functions upon a Group Intelligence Test. Esther Hurley de Weerdt. School and Society, April 9, 1927, 438–440. Results from this study tend to support the conclusion that practice in one function (or type of test) alters another only in so far as the elements are the same.

Intelligence of Rural and Urban Children. John T. Wahlquist. The Elementary School Journal, May, 1927, 682–684. In the Utah Survey rural children's intelligence and educational quotients were lower than those of urban children but their accomplishment ratios were higher.

Achievement Testing

Influence of Grade Progress on Achievement of Subnormal Pupils. Jacob S. Orleans. Teachers College Record, May, 1927, 916–927. Subnormal pupils in the fifth grade achieved more than a year's more work than pupils of the third grade paired with them as to MA and IQ.

The Selective Value of the Examinations of the College Entrance Examination Board. A. D. Whitman. School and Society, April 30, 1927, 522–525. The

author concludes the predictive value of the present College Entrance Examinations is less than that of teachers' marks and suggests the use of more reliable measures which are now available.

Character and Personality

An Investigation of Character Traits in Delinquent Girls. Hazel M. Cushing and G. M. Ruch. The Journal of Applied Psychology, Feb., 1927, 1–7. The results are given of eight tests to 127 delinquent girls and a control group of 79 public school girls.

An Investigation into Some Social Attitudes of High School Pupils. Robert Frederick. School and Society, April 2, 1927, 410–412. An attempt is reported to measure pupils attitudes toward patriotism, race prejudice, international affairs and world peace.

Daydreams: A Cause of Mind Wandering and Inferior Scholarship. Gilbert L. Brown. Journal of Educational Research, April, 1927, 276–279. A study of the daydreams of Normal School students and their effect upon study.

Psychology of School Subjects

Comparison of Two Methods of Teaching Pupils to Apply the Mechanics of Arithmetic to the Solution of Problems. Carleton W. Washburne. The Elementary School Journal, June, 1927, 758–767. The conclusion is reached that the mechanics of arithmetic may be taught with equal efficiency either in practical problems or by themselves.

Summary of Arithmetic Investigations (1926) I. G. T. Buswell. The Elementary School Journal, May, 1927, 685–694. A bibliography of 52 articles is given and the articles containing summaries and studies of drill are summarized.

Summary of Arithmetic Investigations (1926) II. G. T. Buswell. The Elementary School Journal, June, 1927, 731–744. The investigations are summarized under the heads of reading difficulties, methods of teaching, problem solving, time distribution, and curriculum studies.

A Study of Geometric Abilities among Boys and Girls of Equal Mental Abilities. Paul E. Webb. Journal of Educational Research, April, 1927, 256–262. In groups of equal mental ages under 18½ years of ability of the girls in geometry was less than that of the boys. The girls were more variable.

A Study of the Relative Value of Three Methods of Teaching High School Chemistry. H. B. Nash and M. J. W. Phillips. Journal of Educational Research, May, 1927, 371–379. Pupils taught by the lecture and demonstration method gamed more on the test than those who actually worked in the laboratory or those who enjoyed a combination of these last two methods.

Non-school English of High School Students. E. J. Ashbaugh. Journal of Educational Research, May, 1927, 307–313. This is a critical study of 300 letters written by high school pupils and sent by mail to their friends. The per cent of errors is shown for a number of the common types of mistakes.

Teaching Spelling by Column and Context Forms. Paul McKee. Journal of Educational Research, May, 1927, 339–348. Teaching spelling by the common

column method is superior to teaching it in the context of sentences and paragraphs according to two experiments here reported.

Teaching Spelling by Column and Context Forms. Paul McKee. Journal of Educational Research, April, 1927, 246–254. The column method of presenting spelling words proved more efficient than the phrase method in this investigation.

The Relation of Interest to Ability in High School Subjects. W. D. Commins and Theodore B. Shank. The Elementary School Journal, June, 1927, 768–771. The Standard Achievement Test was given to 167 fifth grade children and these pupils were asked to indicate their preference for arithmetic or reading.

College Students' Reactions to Education Courses. George H. Betts. School and Society, April 23, 1927, 494–496. Over 500 men and women in 15 courses in education expressed opinions in regard to classroom procedure, interest, difficulty, etc. in their courses.

Psychology of Childhood

Can Observation Be Trained in School Children? W. H. Winch. Journal of Educational Research, April and May, 1927, 229–238 and 314–326. Carefully controlled experiments with two groups of pupils show that children who are given practice in observation gain 22 per cent and 55 per cent respectively, in ability to observe over similar groups not given practice.

The Compensatory Function of the Movies. Harvey C. Lehman and Paul A. Witty. The Journal of Applied Psychology, Feb., 1927, 33–41. The authors show the popularity of the movies among recreation activities of children and offer an explanation for its high rank.

Some Factors Influencing the Selection of Boys' Chums. Paul Henly Furvey. The Journal of Applied Psychology Feb., 1927, 47–51. Height, weight, developmental age, mental age, chronological age, neighborhood and grade location are the factors studied.

Observations on a Visually Preoccupied Child. June E. Downey. The Journal of Applied Psychology, Feb., 1927, 52–57. The case study of a four year old boy who had particular interest and ability in the discrimination of forms.

Vocational and Educational Guidance

The Predictive Value of Certain Measures of Ability in College Freshmen. Glen U. Cleeton. The Journal of Educational Research, May, 1927, 357–370. A combination of scores on the Thorndike Intelligence Examination and the Iona High School Content Examination gives better prediction of scholarship than either test separately.

Vocational Guidance through School Subjects. Harry D. Kitson. Teachers College Record, May, 1927, 900–915. Numerous examples are given of lesson outlines in conventional school subjects adapted to give an introduction to the vocations.

An Attempt at Predicting Success in the Freshman Year at College. C. W. Odell. School and Society, June 11, 1927, 702–706. This study is based upon the school records, intelligence test scores and answers to a number of questions of nearly 2000 freshmen, formerly seniors in Illinois high schools.

MISCELLANEOUS

Complacency, a Point of View in Education. R. P. Raup. Teachers College Record, May, 1927, 867–883. The author gives a brief explanation of the complacency theory and discusses its application as a means of dealing with certain behavior problems especially in relation to education.

The Correlation Coefficient and Its Prognostic Significance. Clark L. Hull. Journal of Educational Research, May, 1927, 327–338. The prognostic value of measures of various degrees of reliability is discussed. A table and chart are given showing the prognostic value for measures of any reliability.